LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN
numero 57
maggio 1973
anno VI
volume X

Ricognizione e controllo degli armamenti

I satelliti da ricognizione sono il principale strumento mediante il quale sia USA che URSS intendono verificare il vicendevole rispetto degli accordi SALT I. L'importanza di questo e di altri sistemi ai fini dei SALT II

di Ted Greenwood

I due importanti trattati di controllo degli armamenti conclusi nel maggio del 1971 a Mosca, a seguito della prima tornata delle Conversazioni sulla limitazione dell'armamento strategico (SALT I, Strategic Arms Limitation Talks) tra USA e URSS, comprendono entrambi talune clausole relative al problema delle verifiche. L'articolo 12 del trattato sulla limitazione dei sistemi contro i missili balistici (ABM) stabilisce, per esempio, che: « 1. Al fine di garantire il rispetto delle norme del presente trattato, le due parti si serviranno dei mezzi tecnici nazionali di verifica a disposizione di ciascuna... 2. Le due parti si impegnano a non interferire nei mezzi tecnici nazionali di verifica dell'altra parte... 3. Ciascuna delle due parti si impegna a non ricorrere a misure deliberate di mascheramento che impediscano la verifica da parte dei mezzi tecnici nazionali dell'altra parte ai termini del presente trattato... » Il quinto articolo dell'accordo transitorio SALT I, relativo a « certe misure riguardanti la limitazione dell'armamento strategico offensivo », usa un linguaggio praticamente identico per sottolineare i medesimi tre punti.

Quali sono dunque i « mezzi tecnici nazionali di verifica » di cui si fa cenno nei vari documenti? In qual modo questi sistemi vengono impiegati attualmente da ciascuna delle due parti per verificare unilateralmente il rispetto o il mancato rispetto degli accordi SALT I dall'altra parte? Quale incidenza potrebbe avere questa possibilità di reciproca sorveglianza in vista degli ulteriori patti di controllo degli armamenti, in particolare per quanto riguarda gli accordi che potrebbero scaturire dai negoziati SALT II attualmente in corso?

Per tentare di dare una risposta a tali interrogativi, questo articolo dovrà di necessità soffermarsi soprattutto sul tema dei satelliti spia, giacché le principali restrizioni imposte da entrambi i trattati SALT I possono essere (e indubbiamente sono) sostenute dalla sorveglianza condotta a mezzo di elementi rivelatori installati appunto su sistemi orbitanti di fotoricognizione. Le restrizioni primarie previste dai SALT I sono quelle che limitano a 200 le rampe missilistiche ABM in due sole sedi, situate a grande distanza tra loro, che impongono un limite numerico sia ai missili intercontinentali con base terrestre (ICBM, Intercontinental Ballistic Missiles), sia a quelli con base sottomarina (SLBM, Submarine Launched Ballistic Missiles), che regolano la sostituzione degli ICBM e degli SLBM antiquati con i nuovi SLBM e che vietano, infine, il cambio degli ICBM vecchi, quelli leggeri, con i nuovi ICBM pesanti.

Inoltre il trattato sugli ABM bandisce la sperimentazione di nuovi tipi di sistemi ABM e di ogni altro sistema o componente funzionante « con modalità ABM ». La sorveglianza delle attività di questo tipo comporta non soltanto l'impiego dei satelliti spia, bensí anche l'adozione di altri metodi specifici per la verifica dei lanci missilistici. Ecco perché in questo articolo prenderemo in esame in secondo luogo sistemi già piú avanzati, quali i radar con base a terra dell'ultima generazione, i satelliti d'allarme precoce e gli elementi di puntamento e inseguimento installati a bordo di imbarcazioni. Si tratta di sistemi che rivestono particolare importanza nel quadro delle trattative e che non possono essere dimenticati ogniqualvolta si consideri l'eventualità che i SALT II portino all'adozione di restrizioni qualitative significative della missilistica strategica offensiva.

Nonostante la segretezza ufficiale che nasconde quasi interamente alla pubblica opinione i programmi d'osservazione orbitale sia statunitense che sovietica, è sempre possibile proporre alcune considerazioni generali sulle possibilità teoriche di tali sistemi e sulle

Ecco lo schema dei sistemi di ricognizione e sorveglianza utilizzati attualmente dagli USA per verificare il rispetto da parte sovietica delle clausole degli accordi SALT I. Generalmente i satelliti di fotoricognizione americani vengono lanciati (dalla base aeronautica Vandenberg in California) in orbita ellittica quasi polare, con periodo orbitale di circa 90 minuti e perigeo (punto inferiore) dell'ordine di centosessanta chilometri. Il piú recente satellite d'osservazione statunitense (quarta generazione), ufficiosamente denominato Big Bird, riunisce in sé le due funzioni separate della fotografia di sorveglianza di zona e della fotografia a distanza ravvicinata, per cui deve restare in orbita per un periodo di tempo molto piú lungo rispetto ai precedenti satelliti d'osservazione ravvicinata: finora i tempi orbitali sono in media pari a sette settimane. L'orbita di un satellite del genere rimane sostanzialmente fissa nello spazio durante la rotazione terrestre, cosí che all'osservatore al suolo sembra che il satellite si sposti sempre piú a occidente a ogni orbita successiva. In tal modo sotto il percorso orbitale del satellite trascorre la maggior parte della superficie terrestre. I satelliti americani di preallarme vengono invece di solito lanciati in orbite di « parcheggio », quasi equatoriali e quasi sincrone, fino a quote di circa 35 000 chilometri. Due di questi satelliti, lanciati in orbite uguali a forma di 8 sulla stessa longitudine fissa al di sopra dell'oceano Indiano, ma intervallati di 12 ore, sono in grado di fornire la copertura ininterrotta con fotografia all'infrarosso di gran parte dell'Unione Sovietica e dell'intera Cina (*zona in colore chiaro*). Il pregio dei satelliti di preallarme, dal punto di vista del controllo degli armamenti, sta nella loro capacità di sorvegliare i test missilistici. I puntini neri indicano l'ubicazione dei principali centri sperimentali di lancio sovietici; le linee nere rappresentano la traiettoria tipica di un missile sovietico sperimentale a lunga gittata. Sono inoltre illustrati schematicamente due tipi di radar americani usati per seguire i test missilistici russi: trasmittenti oltre-orizzonte (*puntini in colore*) con le corrispondenti riceventi (*cerchietti in colore*), e radar convenzionali con puntamento diretto (*quadratini in colore*).

loro prestazioni pratiche. È noto, per esempio, che, a causa delle proprietà elettromagnetiche fondamentali dell'atmosfera terrestre, i rivelatori installati sui satelliti forniscono prestazioni limitate a tre « finestre spettrali »: le lunghezze d'onda della luce visibile, un'ampia banda dell'infrarosso attorno alla lunghezza d'onda di otto micron, e alcune lunghezze d'onda radar. Con il passar degli anni, con l'aumentare delle dimensioni dei razzi vettori che consentono di mettere in orbita apparati strumentali sempre piú pesanti, è andato aumentando il numero degli elementi di rivelazione, il loro potere risolvente e la quantità di pellicola trasportata. Contemporaneamente si sono fatti sempre piú raffinati i sistemi di comunicazione e di conseguenza la quantità e la qualità dei dati trasmessi.

La misura qualitativa piú utile di ogni sistema di fotoricognizione aerea o orbitale è la risoluzione al suolo, valore equivalente alle dimensioni dell'oggetto piú piccolo che è possibile distinguere con buon contrasto sulla superficie della Terra. Per un apparecchio cinefotografico diretto verso il basso sulla verticale, l'espressione della risoluzione al suolo, G, è data dall'equazione $G = A/300\ FR$, dove A è la quota in piedi dell'aereo o del satellite, F la distanza focale dell'apparecchio (sempre in piedi), R la risoluzione totale della pellicola e dei dispositivi ottici in righe per millimetro, e 300 una costante numerica caratteristica delle unità usate per le altre quantità.

Per citare un esempio recente di fotografia spaziale, ricorderemo che la fotocamera panoramica realizzata dalla Itek Corporation per le ultime tre missioni Apollo, possedeva una distanza focale di due piedi e fotografava la superficie lunare da una quota di circa cento chilometri. Le foto cosí ottenute presentavano una risoluzione al suolo di circa un metro (*si veda l'illustrazione alle pagine 20 e 21*). Se si inseriscono questi dati numerici nella formula precedentemente descritta, si ottiene per il sistema una risoluzione ottica superiore a 180 righe per millimetro.

A titolo di confronto ricorderemo altresí che è stato ufficiosamente comunicato che le fotocamere ad alta risoluzione impiegate nell'ultima generazione di satelliti d'osservazione dispongono di una distanza focale di oltre otto piedi (circa due metri e mezzo). Se si prendono gli otto piedi di distanza focale e le centottanta righe per millimetro di risoluzione ottica, si può calcolare che per il sistema dato, a una quota di 100 miglia (160 000 metri), la risoluzione al suolo è di 1,2 piedi (circa 0,36 metri). In realtà vi è motivo di ritenere che la risoluzione di gran parte dei sistemi avanzati attualmente in esercizio sia anche migliore: già nel 1960 gli esperti del settore parlavano dell'imminente realizzazione di fotocamere dell'areonautica militare statunitense che avrebbero avuto una distanza focale dell'ordine di 20 piedi (6 metri). Inoltre sono ormai disponibili delle pellicole fotografiche con una risoluzione notevolmente superiore a 180 righe per millimetro. Tuttavia è necessario stabilire un compromesso tra il potere risolvente e la sensibilità (o rapidità) della pellicola, giacché per la fotografia orbitale è necessaria una pellicola molto sensibile.

Altro elemento che va preso in considerazione è l'apertura dell'obiettivo fotografico. Gli effetti della diffrazione pongono un limite massimo al potere risolvente di qualsiasi sistema ottico (limite mai raggiungibile in pratica, però). Tale limite è dato dal rapporto $0{,}82d/\lambda F$, in cui d è l'apertura in piedi e λ la lunghezza d'onda in millimetri delle radiazioni alle quali è sensibile il sistema. Per un apparecchio fotografico con distanza focale di otto piedi (2,43 metri), un piede (30 centimetri) di apertura comporterebbe, al limite della diffrazione, un potere risolvente di 180 righe per millimetro, nell'ipotesi che si stia lavorando in luce visibile a una lunghezza d'onda di 0,55 micron.

Per avere la certezza che il sistema ottico non sia limitato dalla diffrazione, si ricorrerebbe in pratica a un'apertura molto piú grande, per cui un'apparecchio con otto piedi di distanza focale e tre piedi di apertura (nell'ipotesi che si usi una pellicola con 180 righe per millimetro di potere risolvente) fornirebbe una risoluzione al suolo di 1,6 piedi da una quota di 160 chilometri. Con la medesima pellicola, un apparecchio avente distanza focale di 20 piedi e apertura di cinque piedi darebbe una risoluzione al suolo di 0,7 piedi. Per conseguire 0,5 piedi di risoluzione al suolo lo stesso apparecchio dovrebbe essere dotato di una pellicola con potere risolvente pari a circa 330 righe per millimetro.

Questi semplici calcoli si basano sull'assunto che il potere risolvente globale del sistema sia inferiore al fattore concomitante piú scarso, che potrebbe essere il potere risolvente al limite della diffrazione per i dispositivi ottici oppure il potere risolvente in righe per la pellicola. Inoltre i risultati cosí ottenuti non tengono conto dei vari limiti operativi, né prevedono la limitazione intrinseca dalla risoluzione al suolo rappresentata dalla dispersione della luce nell'atmosfera, limitazione che, a cento miglia d'altitudine, corrisponde a qualche centimetro. Appare chiaro, nondimeno, che i piú recenti satelliti di fotoricognizione statunitensi siano prossimi al raggiungimento di valori di risoluzione limitati solo dagli effetti atmosferici.

La risoluzione al suolo non è comunque la sola misura che determina il valore qualitativo dell'immagine, in quanto caratterizza l'immagine stessa semplicemente sulla soglia dell'appena percettibile. Sulla percettibilità degli oggetti fotografati influiscono notevolmente il colore e il contrasto luminoso. Non meno importanti sono, sotto questo aspetto, l'angolazione dei raggi solari e il potere riflettente intrinseco dell'oggetto fotografato, giacché questi fattori influiscono sulla luminosità apparente dell'oggetto stesso. In linea di massima in fotografia è sempre piú difficile tentare di identificare un dato oggetto che non semplicemente individuarlo. D'altro canto non è mai impresa molto ardua distinguere le forme allungate, per esempio le strade e i tracciati ferroviari, anche quando le loro dimensioni siano di gran lunga al di sotto della risoluzione calcolata al suolo.

Per il richiamo dallo spazio delle informazioni di fotoricognizione gli USA impiegano attualmente due tecniche diverse. La prima consiste nel rinvio al suolo della pellicola vera e propria per lo sviluppo e la stampa. Questo è un metodo che consente di conservare tutte le informazioni contenute nel film e, quindi, si usa soprattutto per le fotografie ad altissimo potere risolutivo. La pellicola impressionata, espulsa dal satellite, viene restituita al suolo in uno speciale contenitore di rientro che viene intercettato, a metà percorso, da un aereo appositamente attrezzato, mentre compie in paracadute l'ultima parte della discesa (*si veda l'illustrazione a pagina 16*).

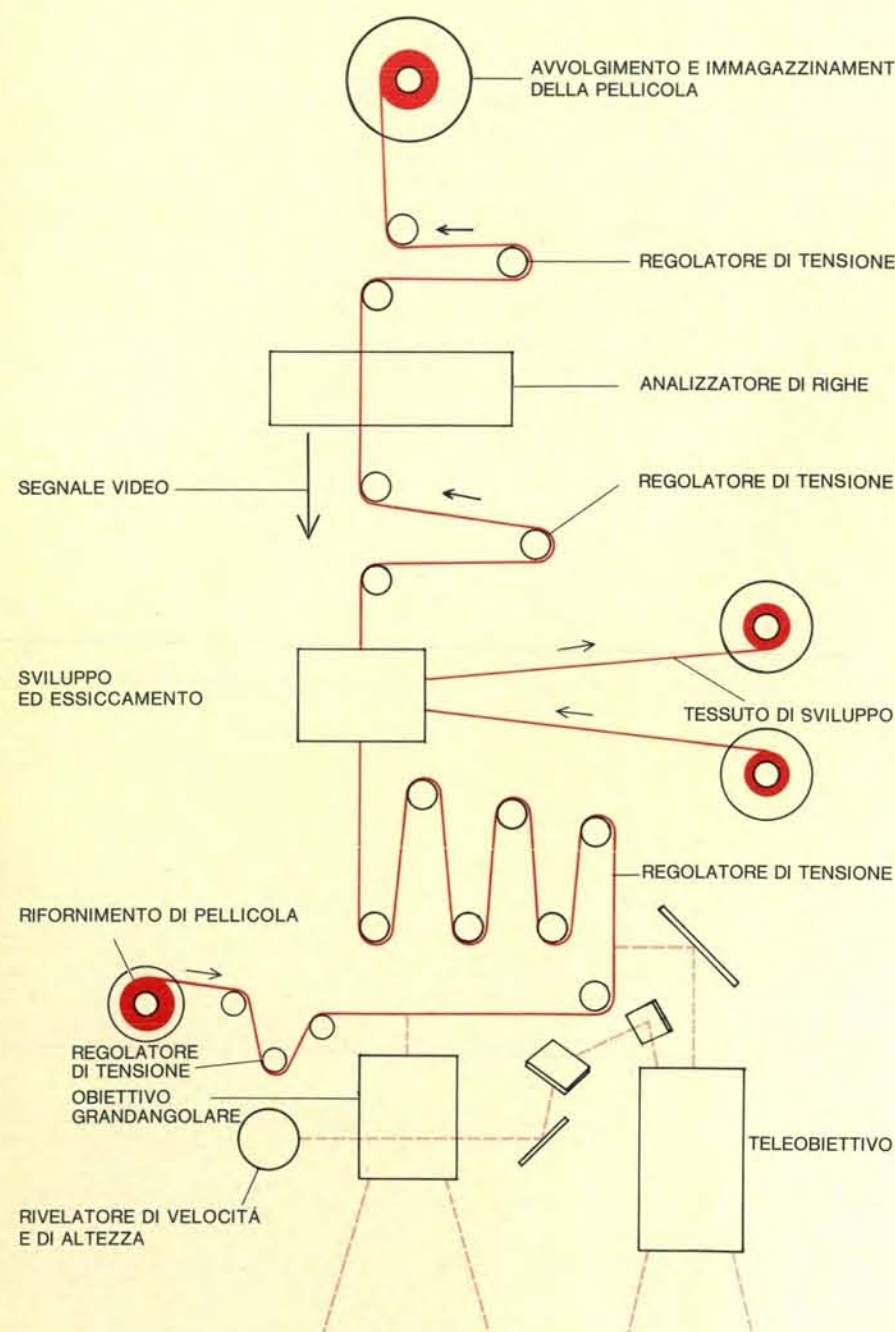
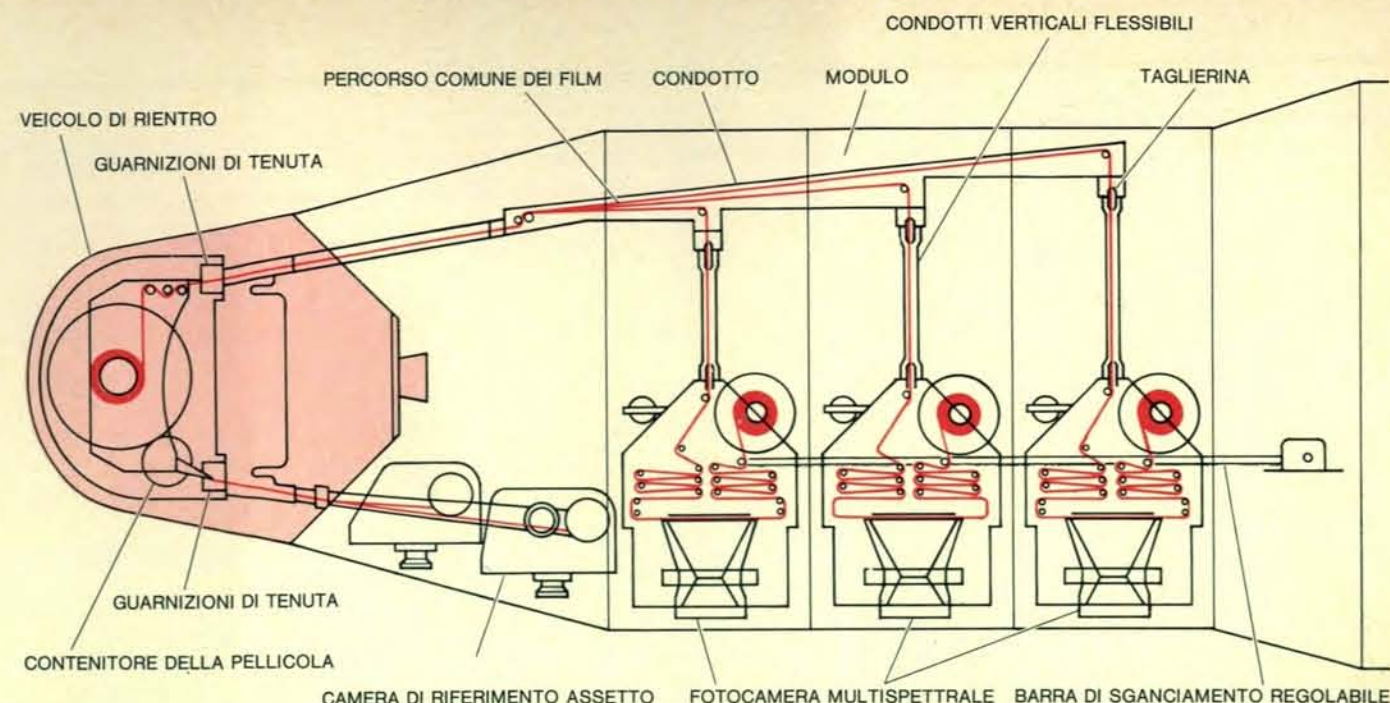
Per le immagini a basso potere risolutivo, la pellicola può essere sviluppata a bordo dello stesso satellite e quindi analizzata da una telecamera o da un impianto laser. Le informazioni cosí raccolte vengono temporaneamente memorizzate e successivamente trasmesse al suolo nel momento in cui il satellite sorvola uno dei tanti osservatori terrestri o navali sparsi sul nostro pianeta. È evidente che, per poter trasmettere tutte le informazioni necessarie nel brevissimo arco di tempo in cui

il satellite resta nel raggio dell'osservatorio, si rende indispensabile un sistema di comunicazioni con un'elevata velocità di trasmissione dei dati.

La pellicola impressionata, una volta rientrata a terra, deve essere sviluppata con il massimo guadagno possibile e quindi interpretata. Con la medesima cura devono essere ricreate e quindi interpretate le immagini trasmesse agli osservatori. Con il potere risolutivo al suolo misurato in pollici e in piedi e con il potere risolutivo della pellicola misurato in centinaia di righe per millimetro, è evidente che ottenere gli ingrandimenti dei fotogrammi e passare alla loro interpretazione richieda un notevole dispendio di mezzi e di tempo. È da credere anzi che ormai, tanto nell'URSS quanto negli USA, la fotointerpretazione rappresenti un'impresa gigantesca. Non deve meravigliare, inoltre, il fatto che per effettuare le fotografie si impieghino strumenti al di sotto del potere risolutivo massimo, poiché solo in questo modo è possibile identificare i bersagli piú interessanti in vista di un'ulteriore ripresa fotografica a maggior potere risolutivo.

A bordo dei satelliti spia si impiegano, oltre agli apparecchi in luce visibile, anche rivelatori all'infrarosso. Poiché alle temperature terrestri tutti i corpi irradiano energia che rientra prevalentemente nella zona infrarossa dello spettro, la fotografia infrarossa non ha bisogno di luce solare per l'illuminazione, per cui risulta estremamente utile per le riprese notturne e nelle regioni polari. (A questo scopo sono stati messi a punto anche sistemi a bassa luminosità che sono probabilmente già in funzione). L'aspetto piú interessante dei rivelatori all'infrarosso sta nella loro capacità di individuare oggetti che, come i silos missilistici, si trovino nel sottosuolo o siano stati mascherati. Nell'immagine infrarossa, giacché il terreno immediatamente circostante all'oggetto che interessa si trova a una temperatura diversa, oppure possiede caratteristiche di emissione diverse da quelle del resto del terreno, il silos sotterraneo balza immediatamente all'occhio. È questo il motivo per cui si impiega sempre piú con i satelliti d'osservazione la fotografia multispettrale, che permette di fotografare contemporaneamente a lunghezze d'onda diverse. È vero, a questo proposito, che la fotografia IR alle grandi lunghezze d'onda richiederebbe un obiettivo molto piú grande per ottenere valori di risolvenza paragonabili a quelli che si ottengono con i sistemi in luce visibile, ma è vero anche che poteri risolutivi cosí elevati non sono necessari per l'assolvimento del compito riservato ai rivelatori all'infrarosso.

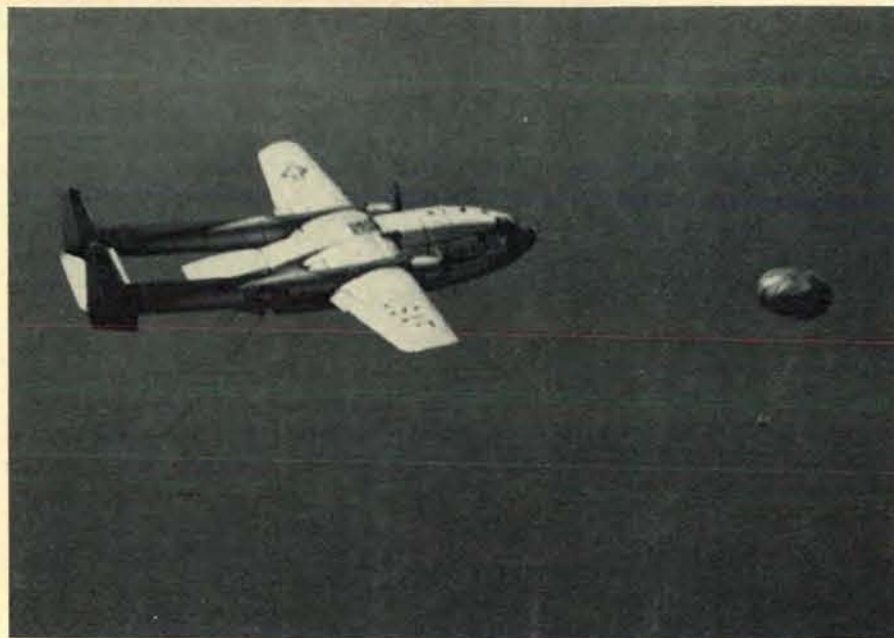
Naturalmente uno dei maggiori ostacoli alla ricognizione orbitale è rappresentato dalle coltri di nubi. È raro, per esempio, che certe località, come la zona di Mosca, siano esenti da vaste formazioni nuvolose. Almeno fino a un certo punto il problema può essere risolto mediante il coordinamento dei lanci dei satelliti d'osservazione con i dati ricevuti dai satelliti meteorologici. Inoltre già i satelliti d'osservazione attualmente in servizio hanno probabilmente la possibilità di mutare orbita per sfruttare ogni varco che si presenti nella coltre di nuvole.

AVVOLGIMENTO E IMMAGAZZINAMENTO DELLA PELLICOLA
REGOLATORE DI TENSIONE
ANALIZZATORE DI RIGHE
SEGNALE VIDEO
REGOLATORE DI TENSIONE
SVILUPPO ED ESSICCAMENTO
TESSUTO DI SVILUPPO
REGOLATORE DI TENSIONE
RIFORNIMENTO DI PELLICOLA
REGOLATORE DI TENSIONE
OBIETTIVO GRANDANGOLARE
TELEOBIETTIVO
RIVELATORE DI VELOCITÀ E DI ALTEZZA

Un sistema di sorveglianza di zona analogo a quello usato per i primi satelliti da ricognizione statunitensi Samos è qui rappresentato mediante lo schema del meccanismo interno di un Lunar Orbiter, il veicolo spaziale impiegato per il rilevamento fotografico della superficie lunare in vista della scelta dei possibili luoghi di allunaggio per l'Apollo. In entrambi i sistemi la pellicola viene impressionata e quindi sviluppata dai dispositivi ottici e di sviluppo contenuti nel sottosistema della fotocamera; quindi le foto sviluppate vengono analizzate da una telecamera che trasforma l'immagine in segnali radio che vengono trasmessi a terra. Tanto per il Lunar Orbiter quanto per i Samos le fotocamere sono state prodotte dalla Eastman Kodak, mentre gli analizzatori televisivi sono stati realizzati dai Laboratori CBS, della Columbia Broadcasting System.

Un sistema di osservazione ravvicinata simile a quello della seconda generazione di satelliti da ricognizione statunitensi Discoverer è qui rappresentato mediante lo schema del meccanismo interno di un prototipo non militare. Questo schema è stato proposto dalla General Electric alla NASA nel 1969, in vista di un suo possibile impiego nel quadro del programma di rilevamento delle risorse terrestri. In entrambi i sistemi la pellicola impressionata passa dal magazzino delle macchine fotografiche in una capsula recuperabile, che viene poi espulsa dal satellite e restituita al suolo per lo sviluppo e la stampa. Il metodo, che conserva sulla pellicola tutte le informazioni possibili, è particolarmente adatto alla fotografia ad alto potere risolutivo. Il sistema a pellicola recuperabile per i satelliti da ricognizione Discoverer è stato realizzato dalla General Electric.

I satelliti spia contribuiscono in vari modi alla verifica degli accordi SALT: fotografando le basi degli esperimenti missilistici possono servire a identificare nuovi sistemi di lancio, a individuare eventuali cambiamenti di procedura operativa indicanti una probabile innovazione di hardware, a sorvegliare l'esecuzione dei programmi sperimentali. Possono osservare gli impianti industriali, compresi gli arsenali per la costruzione di sommergibili e gli stabilimenti per il montaggio dei missili. Si possono ottenere inoltre informazioni intermittenti sui vari aspetti critici delle reti di trasporto e seguire l'avanzamento dei lavori durante la costruzione di silos missilistici, di radar ABM o di rampe di lancio ABM. Le indagini possono essere compiute su scala ancor piú vasta per determinare se siano eventualmente in corso attività che violano gli accordi.

Benché la possibilità di realizzare satelliti terrestri artificiali, con le ovvie implicazioni ai fini della ricognizione militare, fosse stata riconosciuta fin dall'immediato dopoguerra, si dovette attendere la metà degli anni cinquanta prima che fosse assicurata la disponibilità futura dei razzi ausiliari e fosse quindi avviato il primo programma orbitale. Il 16 marzo 1955, sotto gli auspici della CIA, l'areonautica militare statunitense emanò la normativa operativa ufficiale per la realizzazione di un sistema di satelliti strategici. Allo scadere dell'anno previsto per le licitazioni d'appalto, i comandi dell'aeronautica affidarono alla Lockheed Aircraft Corporation l'incarico di mettere a punto un veicolo satellite a propulsione autonoma, in un secondo tempo denominato Agena. Questo veicolo (al pari dei vari modelli successivi) restò per molti anni il cavallo di battaglia del programma d'osservazione orbitale statunitense.

Per iniziare il loro vasto programma di ricognizione strategica, però, gli USA non attesero di avere a disposizione i satelliti: nel 1956 gli U-2 da ricognizione compivano già i loro primi voli nello spazio aereo sovietico. Le informazioni fotografiche cosí ottenute, integrate dai dati radar sugli esperimenti missilistici sovietici, costituirono la base del progressivo ridimensionamento delle stime compiute in precedenza circa l'entità della flotta da bombardamento russa, finendo per convincere inoltre molti osservatori mal informati che l'allestimento dell'arsenale ICBM sovietico procedeva molto piú a rilento del previsto. Nonostante la notevole utilità degli U-2, dal punto di vista della rivelazione gli aerei presentavano due elementi di debolezza. In primo luogo la limitata autonomia dell'aereo, le capacità di resistenza del pilota e la natura provocatoria della missione imponevano un limite rigoroso all'area fotografabile del territorio sovietico. A causa di questo limite, la mancata osservazione di un'estesa rete di ICBM russi da parte degli U-2 non poteva costituire prova certa della scarsa consistenza missilistica sovietica: solo quando i primi satelliti da ricognizione ebbero assicurato una ben piú vasta copertura della terraferma russa, il controspionaggio statunitense poté dirsi certo che non fossero rimasti celati all'osservazione i vari impianti missilistici messi a punto dai sovietici. Il secondo punto debole dell'aereo stava nella vulnerabilità: non passò molto tempo che i russi misero a punto un missile antiaereo in grado di centrare l'U-2 alla quota di ventitremila metri. Allorché nel maggio del 1960 venne abbattuto il pilota Francis Gary Powers, i voli ad alta quota degli U-2, a eccezione di qualche incursione di secondaria importanza, cessarono completamente. Quanto ai satelliti, invece, si sapeva che erano invulnerabili e che tali sarebbero rimasti per il futuro allora prevedibile.

Tra i molti problemi tecnici da risolvere per mettere a punto un satelli-

Il recupero a mezz'aria di una capsula di pellicole sganciata da un satellite da ricognizione tipo Discoverer avviene a mezzo di un aereo appositamente attrezzato mentre la capsula stessa compie, appesa a un paracadute, l'ultima parte della discesa. In questa foto, pubblicata dall'aereonautica militare statunitense nel 1961, si vede un aereo da trasporto C-119 « mentre si avvicina a una capsula Discoverer in un punto imprecisato dell'Oceano Pacifico ». Attualmente, per il recupero delle capsule che vengono espulse dall'ultimo modello di satellite per la ricognizione ravvicinata, si usano i C-130.

te da ricognizione operativo, spiccavano quelli della stabilizzazione e dell'orientamento del veicolo, della progettazione e della produzione di apparecchi fotografici leggeri a grande distanza focale e a grande apertura, nonché il problema del recupero dei dati. La eventualità di impiegare una telecamera, che pure era stata presa in considerazione fin dagli inizi, dovette essere scartata perché era impossibile ottenere, allo stato della tecnologia di allora, l'opportuno potere risolutivo. Nel 1957 fu presa la decisione di seguire due sistemi paralleli: da un lato il recupero diretto della pellicola impressionata, dall'altro lo sviluppo e l'analisi del film a bordo, e successiva trasmissione alle stazioni a terra via radio. Le tecniche per il recupero diretto vennero messe a punto usando come veicoli sperimentali i satelliti Discoverer. Il primo pacchetto di pellicola fu recuperato con successo l'11 agosto 1960 dal Discoverer 13. Forse era questo film o forse il secondo, quello recuperato la settimana seguente dal Discoverer 14, a recare le prime foto orbitali dell'Unione Sovietica. Erano passati tre mesi esatti dall'ultimo sorvolo del territorio russo da parte degli U-2. Il primo satellite di osservazione e radiotrasmissione dei dati funzionante, il Samos II, entrò in orbita polare il 31 gennaio 1961, a una quota variante tra i 480 e i 560 chilometri, con a bordo centotrenta o centottanta chilogrammi di strumentazione.

Quasi certamente furono questi primi successi a liquidare definitivamente il mito del « gap missilistico ». Il 6 febbraio 1961 il segretario americano della difesa Robert S. McNamara, in occasione di una conferenza-stampa improvvisata e informale, annunciò ai giornalisti che un'opportuna indagine aveva rivelato chiaramente l'inesistenza del preteso divario. Un portavoce della Casa Bianca smentí l'annuncio, affermando che l'indagine non era stata ancora ultimata. Ben presto, però, le fonti del controspionaggio cominciarono a ridimensionare i precedenti calcoli sul numero dei missili sovietici. Nel settembre del 1961, a seguito del perfetto recupero di altre capsule Discoverer e nel corso delle analisi sui dati del Samos II, trapelò la notizia che il numero dei missili balistici intercontinentali sovietici poteva essere fissato in quattordici ordigni in tutto.

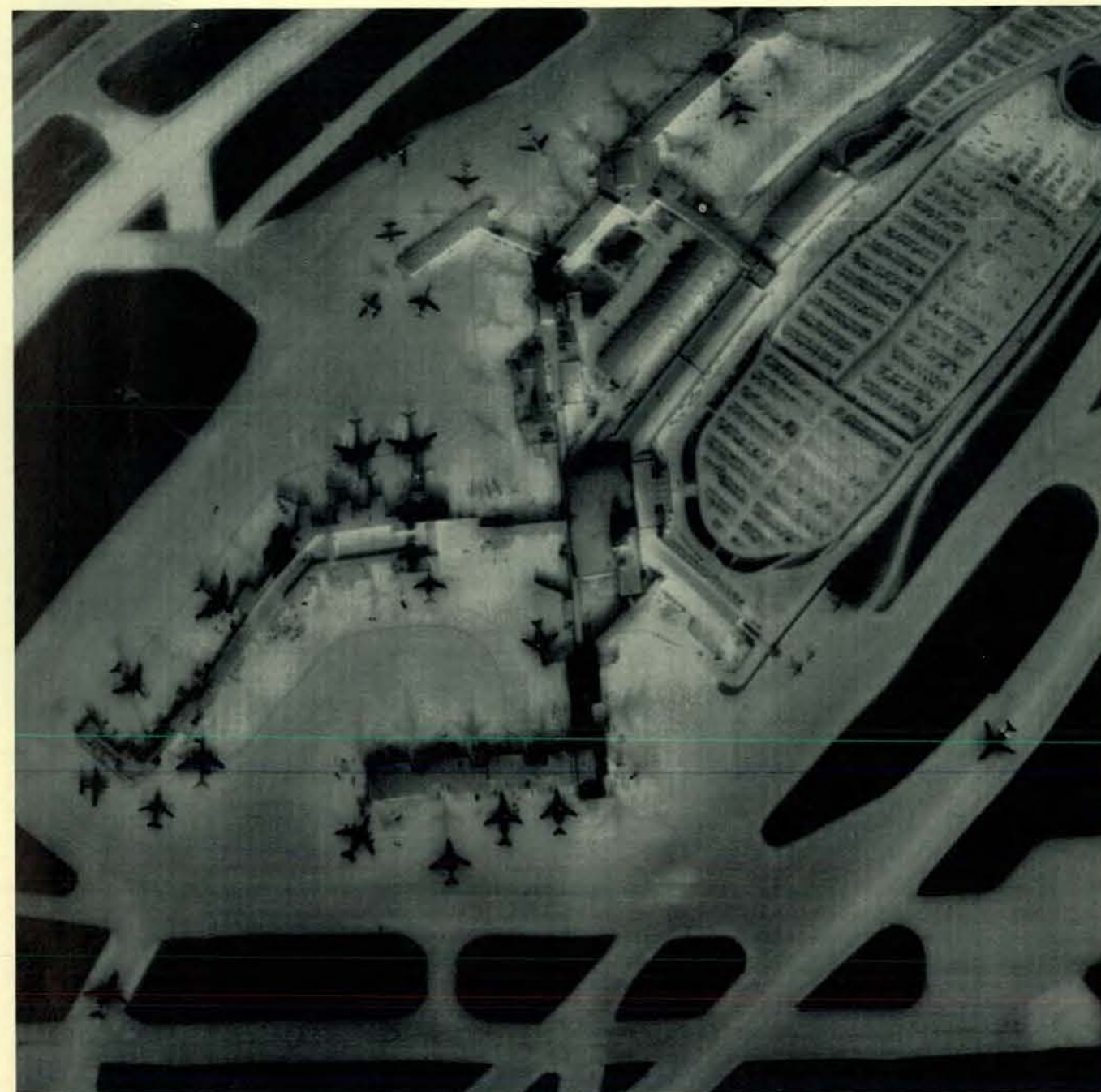
I discendenti dei primi satelliti Samos sono veicoli di osservazione e di radio trasmissione relativamente piccoli, i cui elementi sensibili e le cui caratteristiche orbitali sono destinati a massimizzare l'estensione del campo ricognitivo. Restano in orbita da tre a quattro settimane. Benché il perigeo (punto orbitale inferiore) si trovi all'incirca a centosessanta chilometri dalla superficie terrestre, la loro orbita ellittica riduce enormemente la resistenza aerodinamica, accrescendone quindi la vita utile. Con un'inclinazione da 80 a 92 gradi rispetto all'equatore, questi satelliti assicurano praticamente l'intera ricognizione dell'Unione Sovietica e la completa ricognizione del territorio cinese. In quanto satelliti di « sorveglianza di zona », essi forniscono la copertura a basso potere risolvente di ampie regioni, mentre la trasmissione radio dei dati consente il sollecito recupero delle informazioni ricognitive. Le zone di particolare interesse sono identificate in base a queste informazioni, di modo che in un secondo tempo un altro satellite, questa volta dotato di capsula fotografica recuperabile, possa effettuare fotografie a maggior risolvenza. Agli inizi del 1962 il peso e le dimensioni dell'impianto fotografico erano già stati ridotti di quel tanto che bastava a lanciare i satelliti di sorveglianza di zona mediante il razzo Thor/ /Agena invece che con il ben piú potente Atlas/Agena impiegato prima di allora.

L'uso del Thrust-Augmented-Thor/ /Agena D a partire dal maggio del 1963 sta a indicare l'introduzione di una nuova generazione di satelliti. La maggior forza di spinta di questo razzo ausiliario consentiva al satellite di imbarcare un apparecchio fotografico piú grande nonché un maggior carico utile (comprese le pellicole), conferendogli cosí una maggior durata di esercizio. A partire dalla metà del 1963 questi satelliti vennero lanciati a intervalli mensili, ma fino alla fine del 1965 è capitato spesso che fossero in orbita contemporaneamente due o piú satelliti di sorveglianza di zona e che si verificassero di tanto in tanto dei ritardi rispetto al regolare programma di lancio. Ciò fa pensare che il problema dell'affidabilità richiedesse la sostituzione dei satelliti che si fossero guastati prima del completamento della missione, ma anche che qualche problema sorgesse talvolta a terra. Dal 1966 in poi, invece, questi satelliti restano in orbita ciascuno per circa sei mesi, quasi senza sovrapposizioni nel tempo, per cui è forse lecito fissare ai primi del 1966 la data d'inizio dell'operatività piena e regolare dei veicoli di sorveglianza di zona.

Il primo lancio sperimentale con l'ausilio dell'ancor piú potente Long-Tank-Thrust-Augmented-Thor/Agena risale all'agosto del 1966, mentre l'entrata in servizio regolare del razzo stesso si è avuta nel maggio del 1967. L'avvenimento segna l'avvento di un satellite piú pesante, della terza generazione,

probabilmente munito di un apparecchio fotografico a maggior distanza focale, di una maggior riserva di pellicola, di un impianto ottico a infrarossi e di un nuovo sistema di trasmissione con una maggiore velocità nell'invio dei dati.

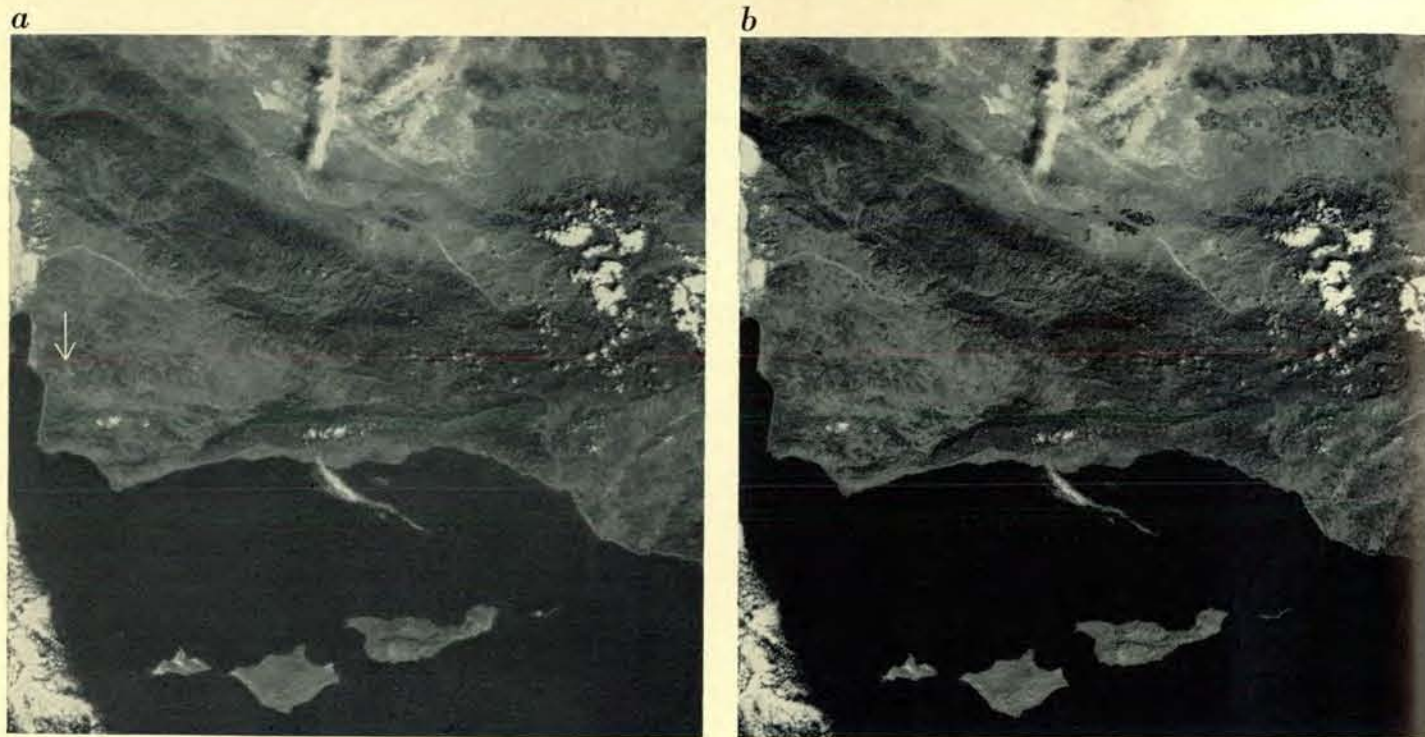
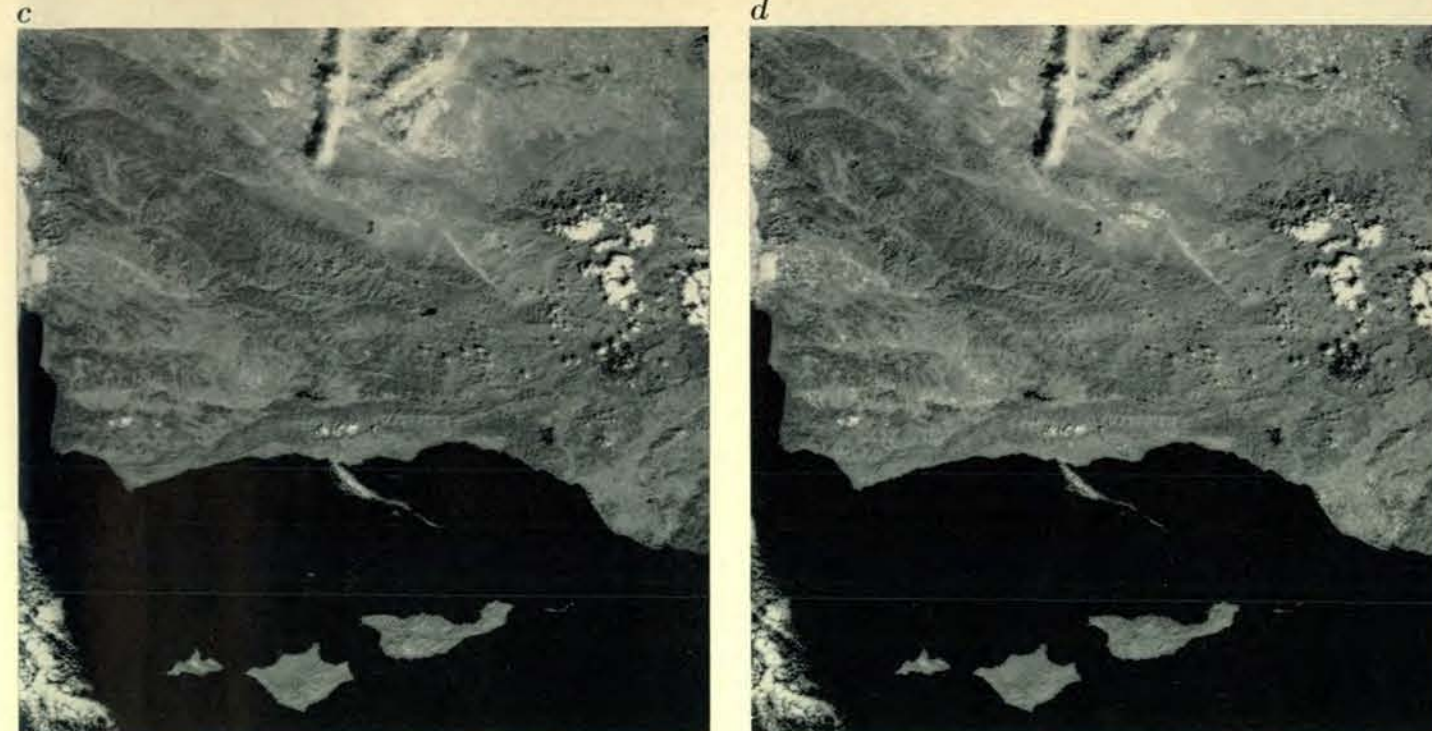
Compito di questi satelliti è l'esplorazione di vaste zone mediante apparecchi sensibili a potere risolutivo medio, nonché la rivelazione di bersagli che meritino un piú attento esame con una maggior risoluzione. Questo esame piú attento viene condotto però da un tipo di satellite diverso.

I discendenti dei satelliti Discoverer a capsula recuperabile sono i satelliti per l'osservazione ravvicinata. Piú pesanti di quelli per la sorveglianza di zona, recano a bordo un apparecchio fotografico con maggior distanza focale e maggior apertura. Operano inoltre secondo un'orbita piú bassa, con il perigeo tipicamente stabile sui centotrenta chilometri allo scopo di migliorare il potere risolvente. Grazie a questo piú potente sistema telescopico, è possibile rifotografare ed esaminare piú attentamente gli obiettivi piú interessanti individuati in precedenza da un satellite di sorveglianza di zona. Per ridurre al minimo le perdite d'informazione conseguenti alla memorizzazione elettronica e alla trasmissione dei dati, i satelliti per l'osservazione ravvicinata rinviano al suolo le loro pellicole racchiuse in una capsula di rientro che viene recuperata a mezz'aria da parte di normali aerei.

Questa foto agli infrarossi di un aereoporto del Texas è stata ripresa da un aereo che volava a seicento metri di quota. Poiché la fotografia infrarossa non ha bisogno dell'illuminazione solare, la si impiega a bordo di satelliti da ricognizione per le riprese notturne o nelle regioni polari. La capacità dei dispositivi e delle pellicole agli infrarossi di intercettare il calore dei gas di scarico dei missili è rivelata dall'immagine abbagliante dello scappamento dell'aviogetto che si trova sulla pista in basso a destra.

a

b

c

d

La fotografia multispettrale, tecnica che permette di scattare contemporaneamente piú fotografie su lunghezze d'onda diverse, è impiegata sempre piú sui satelliti da ricognizione. Le quattro foto-campione riprodotte in queste due pagine sono state eseguite da un analizzatore multispettrale a bordo del satellite per la tecnologia delle risorse terrestri (ERTS 1); in questo caso il dispositivo di scansione delle righe sfrutta uno specchio oscillante per registrare simultaneamente la superficie terrestre che passa sotto il veicolo spaziale su quattro bande spettrali: da 0,5 a 0,6 micron (*a*); da 0,6 a 0,7 micron (*b*); da 0,7 a 0,8 micron (*c*); e da 0,8 a 0,11 micron (*d*). I segnali elettronici prodotti dal sistema televisivo sono successivamente rinviati sulla Terra dove vengono trasformati in immagini in bianco e nero che, a loro volta, possono essere usate per eseguire composizioni cromatiche mediante registrazione sequenziale su pellicola a colori attraverso opportuni filtri. In particolare queste foto sono state qui riprodotte solo per esemplificare la tecnica multispettrale: il loro potere risolutivo al suolo (60 metri circa da una quota di 915 chilometri) non può reggere il confronto con la risoluzione ottenibile con l'attuale generazione di satelliti per la ricognizione ravvicinata (all'incirca trenta centimetri da un'altitudine di centosessanta chilometri). La regione illustrata nelle quattro fotografie è quella circostante alla base aerea statunitense di Vandenberg indicata dalla freccia nella prima foto a sinistra.

Sembra che il primo lancio di un satellite d'osservazione ravvicinata abbia avuto luogo il 26 aprile 1962, allorché un vettore Thor/Agena mise in orbita un satellite dotato di una capsula di recupero contrassegnata con la sigla E-6. Tre giorni dopo furono recuperate tanto la capsula quanto la pellicola. A giudicare dalla frequenza dei lanci sembrerebbe che il programma sia entrato in fase operativa regolare verso la metà del 1963. A motivo della piú breve durata in orbita, i satelliti per l'osservazione ravvicinata presentano esigenze di affidabilità meno imperiose rispetto ai modelli per la sorveglianza di zona, per cui è spiegabile che siano entrati piú rapidamente in fase operativa. Nel 1964 il razzo vettore per l'intero programma era diventato l'Atlas//Agena, che offriva la possibilità di immettere in orbita satelliti anche piú pesanti. Questi veicoli della seconda generazione, dotati di apparecchiature a grande potere risolutivo, venivano lanciati a intervalli di circa un mese e, prima di rinviare la pellicola sulla Terra, restavano in orbita da tre a cinque giorni.

Il Titan 3B cominciò a essere usato per i lanci sperimentali nel luglio del 1966 ed entrò in servizio regolare nell'agosto del 1967. Con l'adozione di questo razzo ancor piú grande dei precedenti, cominciò ad aumentare anche la durata in esercizio dei satelliti per l'osservazione ravvicinata che, entro il 1968, erano già in grado di restare in orbita per un periodo medio di due settimane circa. La circostanza dimostra chiaramente come fosse stato introdotto un satellite di terza generazione dotato di una riserva di pellicola di gran lunga maggiore e della possibilità di allargare l'orbita allo scopo di evitare una disintegrazione termica precoce. Si è detto che il nuovo satellite poteva modificare la sua orbita per sfruttare ogni varco nella coltre di nubi; sembra inoltre verosimile che in quell'occasione furono imbarcati elementi rivelatori di nuovo tipo: i candidati piú probabili dovevano essere i rivelatori agli infrarossi e i dispositivi fotografici multispettrali, che si raccomandavano per la capacità di scoprire e vanificare i mascheramenti nonché, nel caso degli infrarossi, di funzionare anche nell'oscurità. È probabile che la dotazione contenesse anche una macchina fotografica per la cartografia di precisione, destinata a registrare fedelmente l'ubicazione dei vari bersagli in territorio sovietico.

Negli ultimi diciotto mesi circa è stata introdotta, ed è in procinto di entrare in impiego regolare, la quarta generazione, interamente rinnovata, di satelliti da ricognizione. Il nuovo apparecchio, denominato ufficiosamente Big Bird, supera i diecimila chilogrammi di peso, il che significa che è di gran lunga il piú pesante di tutti quelli che l'hanno preceduto. Il veicolo vero e proprio è un missile Agena modificato, del diametro di 3 metri e lungo 15. Il lancio avviene mediante il potente razzo ausiliario Titan 3D.

Il peso e le dimensioni eccezionali del sistema sono dovuti alla fusione in un corpo unico delle funzioni, in precedenza separate, della fotografia di sorveglianza di zona e di quella a distanza ravvicinata. Sembra che il Big Bird imbarchi una fotocamera di sorveglianza di zona della Eastman-Kodak e un gruppo di sviluppo della pellicola e di analisi. I dati che si ottengono vengono trasmessi mediante una nuova antenna allungabile e orientabile di sei metri che, rispetto alle precedenti antenne da un metro e mezzo, avrebbe una potenza sedici volte superiore. In passato ci volevano diversi mesi prima di poter lanciare il satellite per l'osservazione ravvicinata allo scopo di rifotografare una zona di particolare interesse precedentemente individuata dalla fotografia a bassa risolvenza. Altro tempo passava prima di poter disporre delle foto sviluppate e ingrandite. Ora invece è possibile ordinare al Big Bird di puntare il suo apparecchio fotografico ad alta risoluzione (prodotto dalla Perkin-Elmer Corporation) al passaggio successivo: la pellicola impressionata, che sembra abbia un potere risolutivo inferiore ai trenta centimetri da una quota di centosessanta chilometri, torna poi alla base in una delle tante capsule di recupero di cui è provvisto il veicolo. Oggi il tempo d'attesa dovrebbe essersi ridotto quindi a qualche settimana.

La doppia capacità funzionale impone al Big Bird di restare in orbita piú a lungo rispetto ai satelliti spia per l'osservazione ravvicinata delle precedenti generazioni. A questo scopo il veicolo viene inserito in un'orbita piú allargata e al tempo stesso piú ellittica. Le caratteristiche orbitali del primo Big Bird, lanciato il 15 giugno 1971, erano un perigeo di 111 miglia, poco meno di 178 chilometri, e un apogeo di 180 miglia (290 chilometri circa). Per compensare la piú alta quota e per migliorare il potere risolutivo, è stato necessario aumentare, rispetto ai modelli precedenti, sia la distanza focale sia la apertura della fotocamera a grande potere risolutivo. Per assicurare una maggior durata in servizio, il satellite è stato dotato di un razzo propulsore che consente l'ulteriore allargamento dell'orbita e, quindi, impedisce la disintegrazione precoce. I tempi in orbita per i primi tre satelliti della nuova serie sono stati rispettivamente 52, 40 e 68 giorni. Il quarto lancio si è avuto il 10 ottobre 1972.

Certamente a partire dal 1962, ma probabilmente già da prima, gli USA sono in possesso di informazioni dettagliate sul numero e sull'ubicazione dei missili strategici sovietici. Nel 1967 il presidente Johnson annunciava a un convegno di docenti che la ricognizione orbitale valeva dieci volte il denaro che gli Stati Uniti avevano speso complessivamente per le iniziative spaziali. « Io so quanti missili ha il nemico » dichiarò Johnson in quell'occasione, aggiungendo che queste informazioni avevano impedito al paese di nutrire timori altrimenti inevitabili. Dalla fine degli anni sessanta a oggi il Dipartimento della difesa continua a pubblicare regolari informazioni sul livello dell'arsenale sovietico di ICBM, di SLBM e di ABM. Poiché ormai tanto il Congresso quanto l'opinione pubblica si attendono e ricevono queste informazioni nel quadro delle normali attività del dipartimento a giustificazione delle spese di bilancio, i dati pubblicati servono già da qualche anno sia al governo sia ai suoi oppositori per sostenere le rispettive argomentazioni.

Il compito principale assegnato dai SALT I a questi « mezzi tecnici nazionali di verifica » ormai tutelati per trattato, consiste nel controllare i limiti quantitativi imposti al numero di lanciamissili ABM, di grandi radar, di ICBM e di SLBM. Questo è appunto il compito di controllo degli armamenti al quale i satelliti d'osservazione sono piú idonei. In base alle foto orbitali, gli analisti del controspionaggio sono in grado di sorvegliare i lavori di costruzione dei silos missilistici e il trasporto degli ordigni fino alle loro basi d'impiego. Grazie alla fotografia all'infrarosso e multispettrale, è possibile individuare e guardare oltre i vari mascheramenti, nonché sorvegliare le attività notturne. Normale è la sorveglianza dei cantieri navali per osservare i lavori di costruzione di nuovi sommergibili. Analogamente si possono tener d'occhio le istallazioni lanciamissili e radar ABM. Sembra fuor di dubbio che, finché i russi tengono fede all'impegno di « non ricorrere a deliberati provvedimenti di mascheramento per impedire la verifica da parte dei mezzi tecnici nazionali della controparte », sia possibile controllare con un buon grado di fiducia, mediante la ricognizione orbitale, almeno questi aspetti degli accordi SALT I.

Tuttavia il problema di fondo è quel-

Un esempio recente di fotografia spaziale a grande risoluzione è dato da questa immagine di una striscia di superficie lunare ottenuta dal modulo di comando di Apollo 15 dopo che il modulo contenente i due astronauti era già allunato. Per le riprese fotografiche del nostro satellite da una altezza di un centinaio di chilometri è stata impiegata una macchina panoramica con distanza focale di sessanta centimetri. La strisciata completa qui riprodótta è circa un terzo dell'originale. La foto originale ha infatti un potere risolutivo al suolo di circa un metro, con una risoluzione ottica del sistema di oltre 180 righe per millimetro. Il punto di allunaggio dell'Apollo 15 si trova all'interno del quadratino bianco in basso al centro (*si veda l'illustrazione a pag. 22*).

lo di sapere se i sovietici riusciranno prima o poi ad aggirare segretamente le clausole del patto, conseguendo cosí qualche vantaggio politico o militare sugli Stati Uniti. È vero che è praticamente impossibile l'installazione clandestina delle armi proibite, ma non potrebbero i sovietici semplicemente fabbricare altri missili o radar e quindi denunciare il trattato a tempo opportuno, cioè quando fossero in grado di costituire una rete missilistica troppo vasta e potente per dare agli USA la possibilità di reagire tempestivamente? Se si volesse approfondire convenientemente l'argomento bisognerebbe presentare anche una valutazione delle pressioni interne che potrebbero indurre l'URSS a questo tipo di comportamento, uno studio sulle possibili ripercussioni in politica internazionale e infine un'analisi di ciò che i sovietici potrebbero o dovrebbero fare per assicurarsi un certo vantaggio a livello politico o a livello militare.

Considerazioni di questo genere esulano dal nostro discorso, ma si può sempre impostare il problema almeno nelle grandi linee. La sorveglianza ininterrotta da parte dei satelliti d'osservazione delle reti di trasporto sovietiche, delle centrali di produzione dell'energia e degli impianti di fabbricazione dovrebbe garantire agli USA la possibilità di individuare le eventuali attività clandestine in tempo utile per reagire in qualche modo. La riduzione conseguita dal Big Bird dell'intervallo di tempo che passa tra il primo sintomo di attività sospetta rilevato in un fotogramma di sorveglianza di zona e la successiva fotografia ad alto potere risolutivo, serve indubbiamente a limitare il grado di incertezza. È vero che la fotografia non è in grado di penetrare negli edifici chiusi, ma le tecniche all'infrarosso e multispettrali possono spesso dir molto sulle attività che vi si svolgono all'interno, soprattutto qualora attività di interesse particolare fossero chiaramente indicate da inevitabili cambiamenti della procedura operativa.

La coltre nuvolosa resta sempre un grave impedimento all'azione ricognitiva, ma con l'allungamento dei tempi utili in orbita e grazie alla possibilità di modificare il percorso, i nuovi satelliti sono meno vincolati rispetto ai loro predecessori. Ai fini del chiarimento di eventuali ambiguità, possono riuscire utili le tecniche unilaterali di informazione elettronica nonché l'apparecchiatura bilaterale a disposizione della Commissione consultiva permanente (costituita sempre in base agli accordi SALT I). Si può dire insomma che già le dimensioni e la complessità dell'attività industriale indispensabile per costruire e mettere in posizione le moderne armi strategiche, combinandosi con l'ampiezza del campo d'osservazione, con il potere risolutivo e con i dispositivi multispettrali dei satelliti da ricognizione, garantiscano la possibilità di verificare anche unilateralmente, e con un grado elevato di fiducia, il rispetto degli accordi SALT I.

Se un accordo futuro dovesse imporre ai sistemi strategici delle restrizioni a carattere qualitativo, gli USA dovrebbero invece affidarsi, per potenziare la ricognizione orbitale, ad altre tecniche di verifica. Piú avanti esamineremo a questo riguardo l'utilità dei sistemi per il controllo degli esperimenti missilistici. Tuttavia anche a tale livello i satelliti d'osservazione possono svolgere un'importante funzione. Tanto per cominciare, qualsiasi miglioramento d'ordine qualitativo apportato all'armamento missilistico, indipendentemente dalla sua natura (nuovi sistemi di guida, ogive multiple, nuovi radar ABM, nuovi SLBM o ICBM avanzati), deve essere inevitabilmente rispecchiato da un cambiamento delle procedure e delle apparecchiature sperimentali e costruttive. Gli eventuali nuovi elementi di hardware devono essere recapitati nelle rispettive sedi operative, e tutte queste attività possono essere agevolmente seguite dai satelliti.

I sistemi di sorveglianza che potrebbero tornare utili nella verifica dei patti che prevedono restrizioni qualitative ai missili balistici, comprendono radar a puntamento diretto con base a terra, radar con puntamento al di là della linea d'orizzonte, sistemi orbitali e particolari rivelatori su mezzi navali per l'osservazione ai punti terminali. Già nell'estate del 1955 era in funzione a Samsun in Turchia un radar statunitense che seguiva i lanci sperimentali di missili in atto presso la base sovietica di Kapustin Yar, a nordovest del mar Caspio. Grazie a questa sorveglianza il lancio dello Sputnik I nell'ottobre del 1957 e gli esperimenti ICBM russi di quello stesso anno non furono affatto una sorpresa per il controspionaggio statunitense e per i funzionari federali che avevano accesso a questi dati: già sapevano delle grandi capacità sovietiche nel settore. A cavallo tra il 1963 e il 1964 era già installato a Diyarbakir, sempre in Turchia, un radar a grande raggio che visualizzava i missili lanciati dal nuovo centro sperimentale sovietico di Tiuratam, a est del lago di Aral.

Per osservare i test sovietici nella fase finale del volo sono anche disponibili diversi radar fissi con base terrestre, uno dei quali è in funzione sull'isola di Shemya, alla periferia dell'arcipelago delle Aleutine, almeno dal 1959. È un radar capace di seguire e di trasmettere dati sui veicoli di rientro sovietici al momento dell'impatto nella zona sperimentale della penisola di Kamciatka oppure nella regione del Pacifico settentrionale che si trova a nordovest delle Midway. Per i test di piú lunga gittata che si concludono nel Pacifico a sudovest dell'isola Johnston è possibile impiegare diversi radar installati per altri scopi nelle Midway, sull'atollo di Bikini, sull'atollo di Kwajalein e sull'isola Johnston.

A differenza del radar convenzionale, quello oltre-orizzonte, denominato OTH (Over The Horizon), non incontra nel raggio di curvatura terrestre un ostacolo al puntamento. Grazie alla riflessione della ionosfera l'OTH è in grado di spingere la sua ricerca a grandissima distanza, consentendo l'individuazione dei missili immediatamente dopo il lancio. L'OTH a « dispersione avanzata », di recente dotazione, è in grado di rilevare i disturbi ionosferici provocati dal getto ionizzato dei gas espulsi dai motori a reazione. Dal momento che ciascun tipo di missile disturba in modo leggermente diverso la ionosfera, il missile appena avvistato può essere identificato secondo la sua caratteristica « firma » del codice OTH. Nel sistema attualmente in funzione si annoverano tre trasmittenti, una a Formosa, una in Giappone e la terza nelle Filippine. A queste trasmittenti fanno riscontro altrettante riceventi dislocate in Italia, in Germania e in un altro paese europeo. Anche se in origine il sistema doveva servire da rete di allarme precoce in vista di attacchi missilistici massicci, dopo il 1968 è servito a individuare gran parte degli eventi sperimentali noti: sono cosí facilmente intercettabili sugli schermi radar tutti i missili a lunga gittata che partono dai centri sperimentali siti in territorio sovietico.

Di pari passo con lo sviluppo dei satelliti d'osservazione tra il finire degli anni cinquanta e il principio degli anni sessanta, fu portato avanti un programma tanto costoso quanto mal riuscito di messa a punto di un sistema di preallarme orbitale. Per anni il progetto soffrí della scarsa affidabilità dell'hardware e dell'incapacità dei dispositivi all'infrarosso di distinguere tra i pennacchi di scarico dei razzi e la luce solare riflessa dai cirri ad alta quota. Nel 1963 poteva già dirsi acquisita una certa capacità, ma solo di recente è entrato in funzione un sistema ad alto grado di fiducia. I primi due veicoli operativi della rete di preallarme orbitale furono lanciati il 5 maggio 1971 e immessi in orbite quasi sincrone sull'oceano Indiano a circa 65° di longitudine est. Le orbite presentano una inclinazione di dieci gradi per offrire una copertura piú settentrionale rispetto a quella consentita dall'orbita sincrona pura, giacché una lieve ellitticità aumenta il tempo a disposizione per il passaggio sull'emisfero settentrionale. I satelliti sono due per consentire la sorveglianza senza soluzione di continuità. Per quanto la missione principale di questi e di altri satelliti immessi a longitudini diverse consista nel fornire il preallarme in caso di attacco con missili intercontinentali, essi rientrano tuttavia nella nostra trattazione in quanto offrono anche la possibilità di tenere sotto sorveglianza gli esperimenti missilistici sovietici.

Per contribuire alla verifica delle restrizioni degli esperimenti previste dal trattato ABM dei SALT I, sono sempre utilizzabili sia i radar OTH sia i satelliti di preallarme, i quali possono essere impiegati altresí in concomitanza con i radar terrestri a puntamento diretto per segnalare quando e dove abbiano luogo i test missilistici e per fornire utili informazioni sul tipo di esperimento in atto. Pertanto questi sistemi potrebbero essere usati per sorvegliare il rispetto di un eventuale trattato che imponesse restrizioni numeriche o geografiche alla sperimentazione missilistica. Finché il test si svolge entro il raggio d'azione del radar a puntamento diretto, è possibile determinarne il punto d'impatto balistico. Con la osservazione radar si possono altresí determinare talune caratteristiche del veicolo di rientro, quali per esempio le dimensioni e la sagoma. La precisione di queste determinazioni dipenderà dalle caratteristiche di dettaglio sia del radar sia del veicolo di rientro.

Le osservazioni piú utili ai fini della sorveglianza dei test missilistici a grande gittata sono quelle effettuate dai mezzi navali e aerei presenti nella zona ove si verifica l'impatto del veicolo di rientro. Informazioni piú dettagliate sul sistema di rientro si possono ricavare dalle osservazioni fotografiche e radar nei punti terminali: se si confronta il punto d'impatto calcolato con quello osservato e se si conoscono le condizioni meteorologiche locali nel punto d'impatto osservato, non è difficile stimare la modalità del rientro nell'atmosfera. Si può calcolare per esempio se il veicolo di rientro è destinato a sorpassare il punto d'impatto balistico oppure se ne è prevista la caduta « corta. » Dalle immagini radar a breve raggio del veicolo di rientro e dalla sua velocità di attraversamento dell'atmosfera osservata, si possono ricavare utili stime circa il suo peso e il relativo coefficiente balistico (la costante che rappresenta l'efficienza del veicolo nel vincere la resistenza aerodinamica). Qualora sia tentata una manovra terminale comandata a propulsione, gli apparecchi radar e quelli all'infrarosso dovrebbero essere in grado di individuarla. Anche gli esperimenti di testate nucleari multiple, qualora non siano stati rivelati in precedenza mediante altre tecniche, possono essere individuati senza difficoltà presso il punto d'impatto.

Grazie a tante informazioni messe agevolmente a disposizione degli osservatori situati in vicinanza della zona d'impatto, non sorprende che questo tipo di sorveglianza sia ormai in atto da tanti anni. Le unità navali statunitensi sorvegliano i lanci missilistici sovietici nel Pacifico fin dal 1961. Si tratta di osservazioni che non soltanto hanno permesso al controspionaggio di tenersi al corrente sui miglioramenti qualitativi apportati dai sovietici alla tecnologia missilistica, ma hanno anche fornito utili dati sulle « firme » radar dei missili sovietici, costituendo un'impor-

tante fonte d'informazione per la progettazione dei radar ABM. Dell'alta sofisticazione degli attuali strumenti si ebbe chiara dimostrazione il 23 aprile 1970, allorché il Dipartimento americano della difesa distribuí alla stampa una serie di foto di un esperimento sovietico di testata nucleare multipla: si venne a sapere che, dopo l'individuazione radar navale dei veicoli di rientro in arrivo, erano stati inviati sul posto, a scattare le fotografie, alcuni mezzi aeronavali.

La verifica delle eventuali restrizioni agli ICBM che dovessero emergere dai SALT II dovrebbe essere affidata principalmente a questi radar terrestri a puntamento diretto, sistemi OTH, satelliti di preallarme e rivelatori navali. Si impone quindi un esame delle loro possibilità e dei loro limiti quando si voglia offrire una valutazione realistica di quanto è possibile e auspicabile nel quadro dei SALT II.

Una delle soluzioni che sono state proposte al fine di impedire i perfezionamenti qualitativi della missilistica offensiva prevede la limitazione numerica degli esperimenti consentiti annualmente alle parti: si sostiene infatti che, se il numero chiuso fosse sufficientemente modesto, lo stato di concorrenza tra l'hardware esistente e la messa a punto del nuovo basterebbe a scongiurare tutti o quasi i nuovi sviluppi. A parte la questione delle incertezze e delle disparità che potrebbero verificarsi in relazione a questo tipo di restrizioni, è possibile dire, con una certa tranquillità, che non dovrebbe essere difficile la verifica della limitazione in sé. Il conteggio dei lanci missilistici è proprio quello che sanno far meglio i satelliti di preallarme, i radar OTH e i radar fissi terrestri. Se scopo dell'accordo è però la limitazione del miglioramento qualitativo, un utile allargamento del concetto potrebbe essere il patto che tutti i test autorizzati siano condotti all'interno di corridoi di traiettoria predesignati e dietro debito preannuncio. L'estensione faciliterebbe il compito di garantire che, limitazioni numeriche a parte, non siano stati apportati miglioramenti qualitativi.

Per quanto riguarda le limitazioni pratiche d'ordine qualitativo, la questione della verifica si fa molto più complessa, ma un certo ottimismo dovrebbe nascere dalla lungimiranza delle restrizioni allo sviluppo previste dal trattato ABM. In linea di massima dovrebbe essere possibile verificare con elevato grado di fiducia, soprattutto se integrata nelle limitazioni suggerite in precedenza, la proibizione d'ogni tipo di nuovo razzo ausiliario, di configurazione bifrontale, di veicolo di rientro: dal momento che le varie « firme » di identificazione dei sistemi attuali sono già note o lo saranno in breve tempo, ogni nuovo elemento di hardware sarebbe identificato grazie alla novità stessa della « firma ». Se tanto l'Unione Sovietica quanto gli Stati Uniti fossero disposti ad accettare certe asimmetrie tecniche, un punto fermo di questo genere potrebbe risultare utilissimo e fors'anche realizzabile in pratica.

Si possono considerare altri divieti a carattere meno restrittivo. Per esempio, giacché le manovre nei punti terminali sfuggono difficilmente alla ricognizione, sia direttamente mediante gli apparecchi di rivelazione infrarossa e radar, sia indirettamente mediante il confronto tra traiettoria calcolata e traiettoria osservata, tali manovre potrebbero rientrare nelle proibizioni. Un divieto di questo genere escluderebbe in partenza ogni strumento di guida in fase terminale, in quanto dispositivi del genere sarebbero inutili visto il divieto delle manovre correttive.

Altra possibilità, sia pure ben più difficilmente verificabile, potrebbe esser quella di vietare il perfezionamento del grado di precisione: l'osservatore è sempre in grado di prevedere dove va ad atterrare un veicolo di rientro e, probabilmente, è anche in grado di stabilire se il veicolo ha subito una manovra terminale intenzionale o se ha sfruttato una particolare conformazione aerodinamica per giungere sul bersaglio, ma in base alla semplice osservazione nessuno saprà mai quale fosse in origine il bersaglio esatto desiderato. Le informazioni sulla precisione dei missili devono quindi derivare da deduzioni di secondo ordine. Un modo per stabilire la precisione dei sistemi di guida, ma non quella delle tecniche di rientro, consiste nel valutare la precisione apparente dei sistemi di guida inerziale adottati per le missioni spaziali. È da presumere che la tecnologia disponibile per i voli nello spazio sia anche utile per i missili intercontinentali. Gli USA dispongono di una capace rete mondiale di radar inseriti nel sistema di rivelazione e di inseguimento spaziale: questi impianti possono essere usati appunto per trarre le debite conclusioni sui missili guidati. Nondimeno non pare che il tentativo di limitare la precisione mediante il vincolo imposto ai sistemi di guida inerziale costituisca la giusta soluzione. Meglio sarebbe ricorrere al divieto di dotare i veicoli di rientro di dispositivi per la manovra terminale e di elevati coefficienti balistici, in modo da impedire in via indiretta ogni miglioramento della precisione.

Altro tema preoccupante all'ordine del giorno dei colloqui SALT II è quello degli esperimenti con testate nucleari multiple. La presenza di più ogive belliche potrebbe essere determinata a mezzo dei radar terminali, dei radar a bordo di unità navali ed eventualmente di altri sistemi. Ma non sarebbe possibile sperimentare un ordigno a testata multipla dotato di una sola ogiva per volta, per sfuggire cosí all'individuazione? La risposta completa al quesito dipende necessariamente sia dai dettagli strutturali dell'hardware sia dal grado di affidabilità del sistema in questione; si può invece formulare una risposta parziale nel senso che, in un mondo caratterizzato dalla deterrenza reciproca, preoccupazione prima delle parti in campo è che l'avversario non riesca ad acquistare una capacità tale da rendere possibile un attacco preventivo a sorpresa. Ebbene, sembra poco verosimile che una tale capacità sia realizzabile e ancor meno che la si possa portare a un buon grado di affidabilità in mancanza di un elevato numero di esperimenti a sistema completo. Resta comunque il problema di tradurre la probabilità che l'Unione sovietica non riesca a creare un tale sistema a grande affidabilità, qualora fosse vincolata da un trattato di controllo degli armamenti, nella certezza da parte degli Stati Uniti che un sistema del genere non sia stato effettivamente messo a punto.

A questo proposito giova a poco, in quanto dipendente dalla natura del sistema, la distinzione tra MRV (veicoli multipli di rientro) e MIRV (veicoli di rientro multipli orientabili indipendentemente). Per una separazione minima tra i vari veicoli di rientro, infatti, potrebbero essere escogitati sistemi dotati sia di elementi di separazione meccanica, sia di strumenti per la guida indipendente. Ogni tentativo di distinzione tra un sistema del genere e un altro sarebbe quindi difficile e scarsamente valido. Inoltre è da presumere che esista la possibilità di programmare un sistema di separazione ampia in modo da avere una separazione minima e procedere successivamente alla sperimentazione al riparo d'ogni rischio d'identificazione. Appare molto difficile, insomma, imporre limitazioni verificabili ai sistemi a testata nucleare multipla, se non nel senso del divieto totale o di un congelamento dei sistemi attuali.

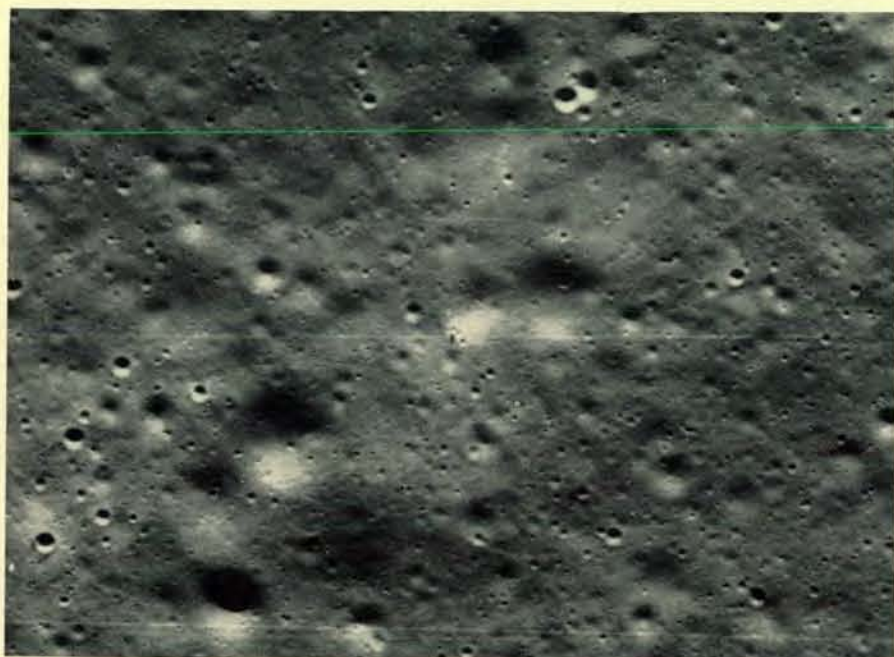
Nel presente articolo ci siamo occupati dei mezzi tecnici nazionali di verifica in quanto si tratta delle tecniche previste e tutelate in base agli accordi SALT I. Non se ne deve però trarre la conclusione che tali mezzi tecnici di verifica siano i soli a disposizione. Tra le altre tecniche di verifica nazionali (o unilaterali, rispetto a quelle « sul posto » o di cooperazione) ricorderemo l'analisi economica, l'attività diplomatica, l'analisi di documenti e discorsi, i colloqui con turisti e partecipanti a congressi scientifici e, infine, lo spionaggio. Queste fonti d'informazione, benché sia ingiusto negarne la grande utilità, presentano sempre lo svantaggio di basarsi su deduzioni, mentre la ricognizione e la sorveglianza dipendono essenzialmente dalle caratteristiche fisiche dei rivelatori elettromagnetici, per cui assicurano una minore ambiguità informativa e un più ampio campo conoscitivo.

La presente trattazione non può naturalmente pretendere di aver esaurito il tema dei mezzi tecnici nazionali di verifica. Omissioni notevoli sono state, per mancanza di sufficienti informazioni non segrete, le unità navali, gli aerei, i satelliti e le riceventi con base terrestre, impiegati per l'informazione militare elettronica e delle comunicazioni. Le informazioni raccolte da questi mezzi, quali le frequenze di trasmissione radar e la natura delle reti di comunicazione sovietica e cinese, potrebbero assumere ai fini militari una rilevanza di gran lunga superiore a quella della verifica degli accordi SALT. Nella misura in cui è possibile però intercettare e tradurre in chiaro le comunicazioni, potrebbe darsi che il controspionaggio sia in grado di passare dalla valutazione delle capacità tecniche a quella di veri e propri sintomi di intenzione politica.

La conclusione più importante che si può trarre dall'analisi qui presentata è che, grazie ai satelliti d'osservazione e ai sistemi di sorveglianza dei test missilistici, gli USA sono in grado di verificare il rispetto da parte sovietica del trattato ABM SALT I e dell'accordo transitorio, con grande attendibilità. Anche se nell'articolo ci siamo occupati essenzialmente dell'arsenale statunitense, va ricordato che anche i russi dispongono di analoghi sistemi d'osservazione orbitale e che sono quindi in grado a loro volta di fidarsi del rispetto dei trattati da parte americana. Ormai i reciproci vantaggi della verifica vicendevole mediante i mezzi tecnici nazionali sono formalmente riconosciuti nel concordato di non interferenza e di non mascheramento che fa parte degli accordi già firmati.

La conclusione della verificabilità non dipende però dalla misura in cui i sovietici collaboreranno al non mascheramento: anche se è molto più facile sorvegliare una nazione mediante i mezzi della cooperazione, non si può escludere l'eventualità di frode sul trattato del controllo degli armamenti. E lo scopo dei sistemi di ricognizione è di individuare o scongiurare tali frodi.

Gli altri importanti negoziati sul controllo degli armamenti, quelli dei SALT II, sono in corso da alcuni mesi. È molto diffusa la speranza che in questa tornata delle conversazioni l'accordo transitorio SALT I possa essere trasformato in una serie di limitazioni numeriche permanenti dei sistemi offensivi. La verifica del rispetto di un simile accordo presenta gli stessi problemi affrontati dai SALT I, problemi che potrebbero essere parimenti risolti con grande fiducia mediante i satelliti d'osservazione e i sistemi di sorveglianza dei test missilistici. Si potrebbe addirittura verificare la proibizione di sistemi terrestri mobili, benché un'eventuale limitazione numerica che non sia lo zero assoluto possa rivelarsi di difficile controllo a causa della natura intermittente della ricognizione orbitale.

È stata inoltre espressa la speranza che i SALT II siano in grado di bloccare la corsa qualitativa agli armamenti missilistici offensivi. A questo proposito si potrebbe istituire un promettente pacchetto di divieti verificabili, compresa la proibizione totale degli esperimenti di nuovi razzi vettori per ICBM e di nuovi sistemi di rientro. Dal momento che ciascun tipo di razzo o di sistema di rientro presenta una serie di caratteristiche uniche e a sé stanti, non sarebbe difficile l'individuazione delle eventuali novità. Sarebbe meglio se il divieto fosse accompagnato dalla limitazione numerica degli esperimenti consentiti annualmente e dalla clausola che i test consentiti siano debitamente preannunciati e condotti all'interno di corridoi di traiettoria prestabiliti. La prima limitazione ridurrebbe le occasioni di frode, incentivando il rispetto dei patti, dal momento che il nuovo hardware potrebbe essere collaudato solo rinunciando alla possibilità di sperimentare l'hardware già esistente. La seconda limitazione agevolerebbe il compito di verificare il divieto contro il nuovo hardware. Tanto l'una quanto l'altra proibizione potrebbe essere verificata a sua volta, giacché con gli attuali sistemi di sorveglianza è praticamente impossibile condurre clandestinamente attività sperimentali non autorizzate.

Particolare ingrandito del luogo di allunaggio dell'Apollo 15, ricavato dalla strisciata panoramica riprodotta nelle due pagine precedenti: è chiaramente visibile il modulo lunare (*perfettamente al centro*). La risoluzione di questa riproduzione, ottenuta da negativi di terza generazione, è notevolmente più scarsa, come è ovvio, di quella dell'originale, in cui sono visibili invece persino i singoli esperimenti eseguiti dagli astronauti. In altre foto ottenute con questo sistema si vedono anche le tracce della « jeep lunare » usata dagli esploratori. Eppure la risoluzione ottica intrinseca del sistema è ancora di gran lunga inferiore a quella delle più progredite fotocamere per la ricognizione orbitale, che forse riescono a ottenere una risoluzione limitata soltanto dagli effetti atmosferici.

La sicurezza dell'automobile

Le statistiche relative agli incidenti d'auto fanno pensare che sia opportuno modificare la struttura delle automobili. Tra queste modifiche le più importanti sono i frontali deformabili e gli abitacoli rinforzati

di Patrick M. Miller

Può un'automobile essere sostanzialmente piú sicura? Considerate le attuali condizioni delle strade, la varietà dei gusti del pubblico in fatto di automobili e la vasta gamma di capacità e di abitudini di guida, si può progettare un'automobile che riduca l'elevato pedaggio di morti e di feriti che si paga sulle strade fornendo, in caso di collisione, un maggior grado di protezione al guidatore e ai passeggeri? Le indagini compiute nell'ambito del programma « ricerche fondamentali sulla sicurezza in caso d'urto » organizzato dalla National Highway Traffic Safety Administration (NHTSA) indicano come diverse soluzioni strutturali attualmente in fase di sperimentazione sembrino promettenti. Parlerò della valutazione attuale della situazione sulla base degli esperimenti condotti dal Cornell Aeronautical Laboratory (ora Calspan Corporation) che da molti anni compie ricerche sugli incidenti automobilistici e sulle misure di sicurezza.

La sicurezza dell'automobile, nonostante abbia costituito un problema fino dal momento della invenzione di questa, non è stata ufficialmente considerata problema nazionale fino al 1966 quando il Congresso degli USA, in seguito all'aumento allarmante del numero degli incidenti stradali mortali, emanò l'atto relativo alla sicurezza del traffico e delle automobili. Il numero di morti per 100 milioni di chilometri di guida si manteneva pressoché stabile, ma poiché era in continuo aumento il numero degli americani che si mettevano per strada, la quantità totale di incidenti con morti e feriti saliva rapidamente. Il National Safety Council valuta che i totali annui siano attualmente di 16 milioni di incidenti con 55 000 morti e circa due milioni di feriti con invalidità temporanea.

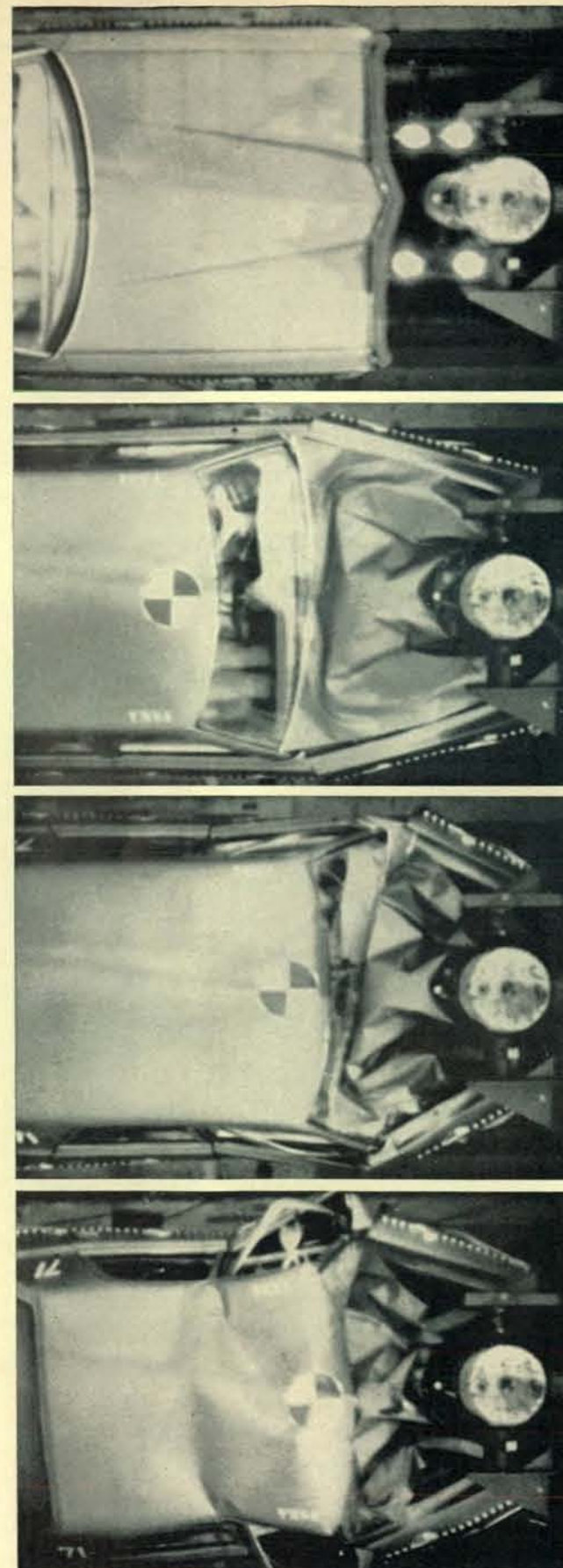
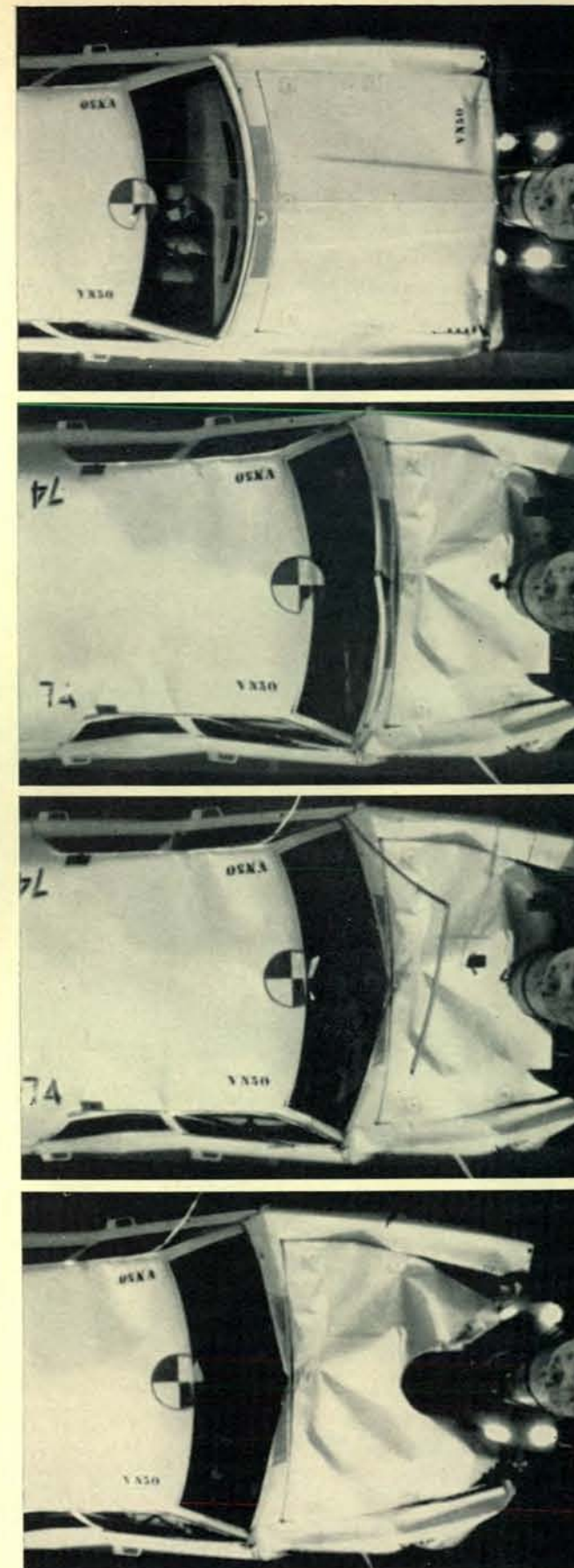
La NHTSA ha istituito diversi programmi, tra i quali alcuni riguardano i guidatori e l'assistenza alle vittime degli incidenti; il maggior sforzo però è accentrato sull'automobile stessa imponendo ai costruttori di auto misure e dispositivi di sicurezza nuovi. Il problema però si rivela molto piú complesso di quanto si potrebbe pensare.

Nel progettare un sistema di sicurezza si deve iniziare col definire esattamente ciò da cui intendiamo difenderci. La NHTSA ha assunto come dato di base che la sicurezza di un autoveicolo in caso d'urto va valutata in termini di capacità di proteggere da danni i viaggiatori e i pedoni, non in termini di costo del danno prodotto al veicolo stesso. Questo però è solo uno dei tanti punti che vanno precisati per giungere a un progetto appropriato.

Per esempio, si deve considerare non solo la protezione dei viaggiatori di una data automobile ma anche la protezione degli occupanti di un'altra automobile con la quale la prima può entrare in collisione. Un frontale robusto e pesante dotato di un'alta capacità di assorbire energia in caso di urto potrebbe dar luogo a un veicolo offensivo che, mentre da un lato protegge i propri occupanti, dall'altro potrebbe costituire un rischio per tutte le altre automobili. Quindi, per salvare delle vite umane un veicolo sicuro deve adeguarsi, in linea generale, alla vasta gamma di dimensioni, pesi e prestazioni dei 110 milioni di autoveicoli attualmente in circolazione sulle strade degli USA.

Un altro fattore di complicazione da considerare è la netta differenza esistente tra i tipi di incidenti che avvengono nelle zone extraurbane e quelli che avvengono nelle zone urbane. Per esempio, piú dei due terzi degli incidenti mortali avvengono nelle zone extraurbane (in questi incidenti la velocità media d'urto di un'automobile è approssimativamente di 80 chilometri l'ora); d'altro canto alle aree urbane con elevata densità di popolazione vanno attribuiti quasi i due terzi degli incidenti automobilistici con danni non mortali alle persone. Gli incidenti in cui si ha il ribaltamento dell'auto sono molto piú comuni nelle zone extraurbane che in città; al contrario gli incidenti in cui un'automobile si scontra con il fianco di un'altra, data la densità degli incroci, sono piú comuni in città.

Vi è ancora un'altra considerazione da fare ed è la necessità di assicurare una migliore protezione ai pedoni. Nel 1971, circa un quinto (10 600) delle persone uccise dalle automobili è rappresentato da pedoni e circa 150 000 sono stati i pedoni feriti. Per i pedoni il rischio costituito dalle automobili è naturalmente molto piú elevato nelle zone urbane che non in quelle extraurbane.

Costruire un'automobile che lasci un pedone illeso dopo averlo urtato è ovviamente un problema quasi insolubile. Sono possibili però alcuni miglioramenti che tendano a ridurre la gravità dell'urto, per esempio l'eliminazione delle sporgenze che potrebbero trattenere o agganciare il pedone e la diminuzione della rigidità delle strutture del cofano e della mascherina del radiatore. Quando un pedone viene investito da un autoveicolo subisce generalmente due urti: il primo direttamente contro il veicolo investitore, e il secondo di rimbalzo contro il manto stradale. Non sembra che ci siano soluzioni per evitare o mitigare il secondo urto. Tutto considerato è chiaro che gli sforzi per proteggere i pedoni debbano essere diretti soprattutto allo studio di sistemi di controllo del traffico che separino le automobili dai pedoni.

Esperimenti d'urto su un'automobile di tipo convenzionale (*a sinistra*) e un'automobile a struttura modificata (*a destra*) sono stati compiuti nell'ambito delle ricerche del Cornell Aeronautical Laboratory. Le fotografie sono state fatte con una cinepresa installata sopra la zona in cui avveniva l'urto. Le automobili venivano lanciate lungo una pista in modo che si scontrassero con un pesante palo d'acciaio a una velocità di 95 chilometri l'ora. L'automobile modificata, che aveva il frontale e il telaio rinforzati e un dispositivo che impediva al motore di penetrare nell'abitacolo, ha subito danni apprezzabilmente minori.

Per quanto riguarda le diverse condizioni di guida in città e fuori città, è opportuno considerare la possibilità di sviluppare diversi progetti di sicurezza, uno per le automobili destinate prevalentemente alla circolazione urbana, l'altro per le automobili usate prevalentemente nell'ambiente extraurbano. Può darsi che le ricerche a questo proposito siano compensate da una riduzione del numero degli incidenti.

Per il momento, però, le ricerche sono concentrate sui problemi e sulle soluzioni a cui si deve dare la priorità assoluta. In un campo cosí complesso e cosí poco esplorato, è piú saggio cominciare con lo stabilire una base razionale su cui sviluppare le conoscenze in maniera graduale. Per capire il problema e saggiare le possibili soluzioni, le informazioni piú utili di cui disponiamo ci vengono dai dati statistici sugli incidenti automobilistici. I campioni che si possono trarre da questi dati ci forniscono suggerimenti per soluzioni specifiche atte a rendere le automobili piú sicure in caso d'urto.

Si sono trovati campioni significativi in un'analisi relativa a circa 25 000 incidenti con feriti, avvenuti prevalentemente in zone extraurbane, registrati nell'ambito del progetto di ricerca condotto dalla Calspan Corporation sulle ferite provocate da scontri automobilistici. Gli incidenti sono stati classificati secondo l'angolo di collisione (scontro frontale, posteriore o laterale) con una categoria a parte riguardante gli incidenti in cui l'autoveicolo si era ribaltato in seguito o no a collisione. La scoperta piú interessante tratta dall'analisi di queste statistiche è stata la dimostrazione che nonostante le collisioni frontali siano gli incidenti piú frequenti, le collisioni laterali e i ribaltamenti sono di gran lunga i piú pericolosi, come risulta sia dall'entità dei danni subiti dall'automobile sia dall'incidenza di ferite gravi o mortali subite dal conducente. Anche se gli scontri laterali costituiscono solo il 13 % di tutti gli incidenti, il 23 % di essi ha avuto come risultato danni ingenti alla vettura e il 16 % ferite pericolose o mortali per il conducente. Analogamente, dei casi di ribaltamento (circa il 21 % di tutti gli incidenti), il 23 % ha comportato gravi danni alla vettura e il 17 % ferite pericolose o fatali per i conducenti. D'altro canto, sebbene le collisioni frontali costituiscano il tipo di incidente piú comune (59 %), solo nel 9,5 % vi sono stati gravi danni e nel 9,6 % si è avuta la morte o quanto meno il grave ferimento del conducente (*si veda l'illustrazione a pag. 33*). Solo una piccola percentuale dei 25 000 incidenti cruenti (4,5 %) è del tipo in cui il veicolo viene tamponato, e questo indica come la protezione della parte posteriore dell'automobile non richieda la stessa attenzione che viene data invece alla protezione della parte anteriore e delle fiancate.

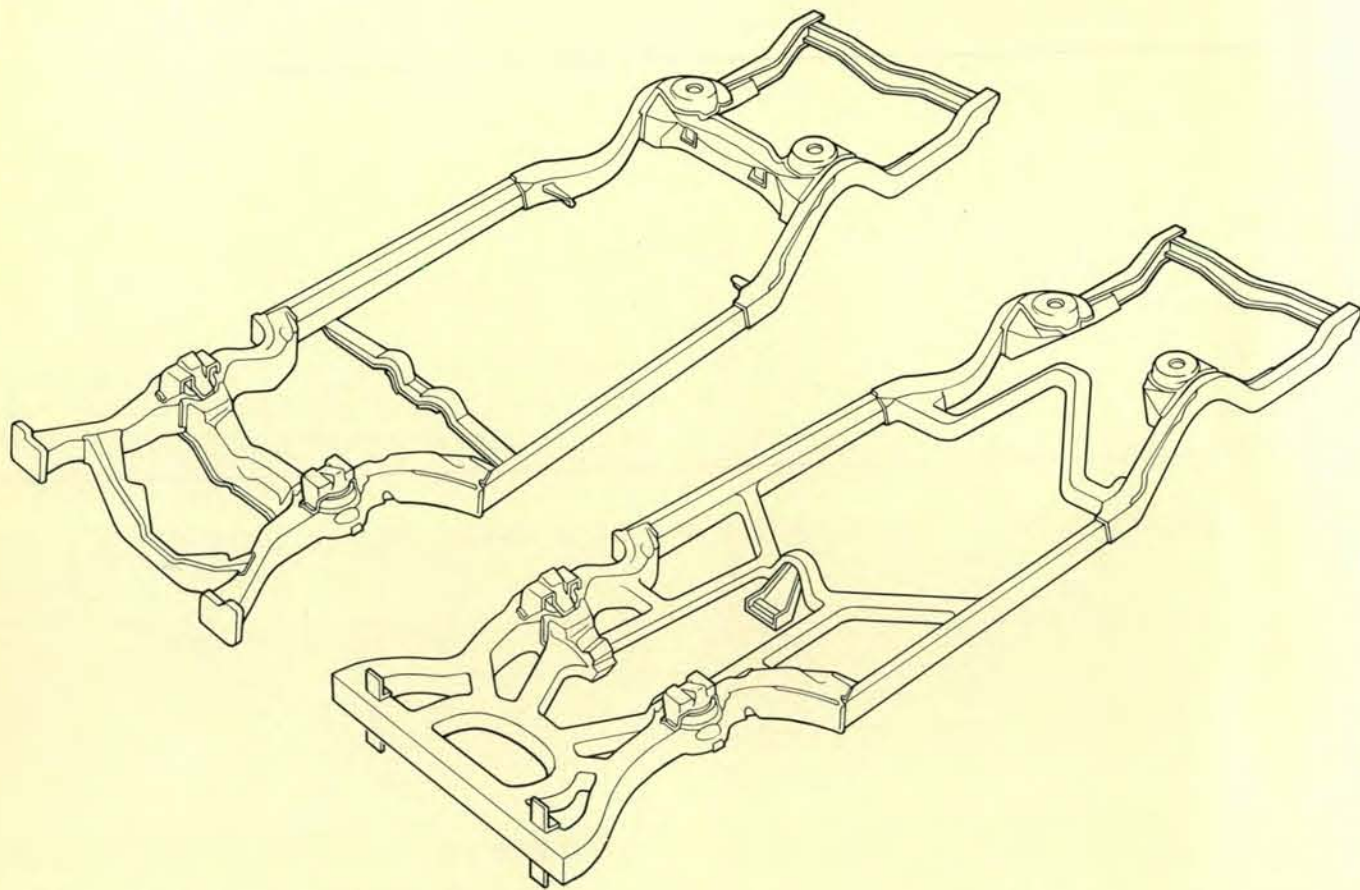
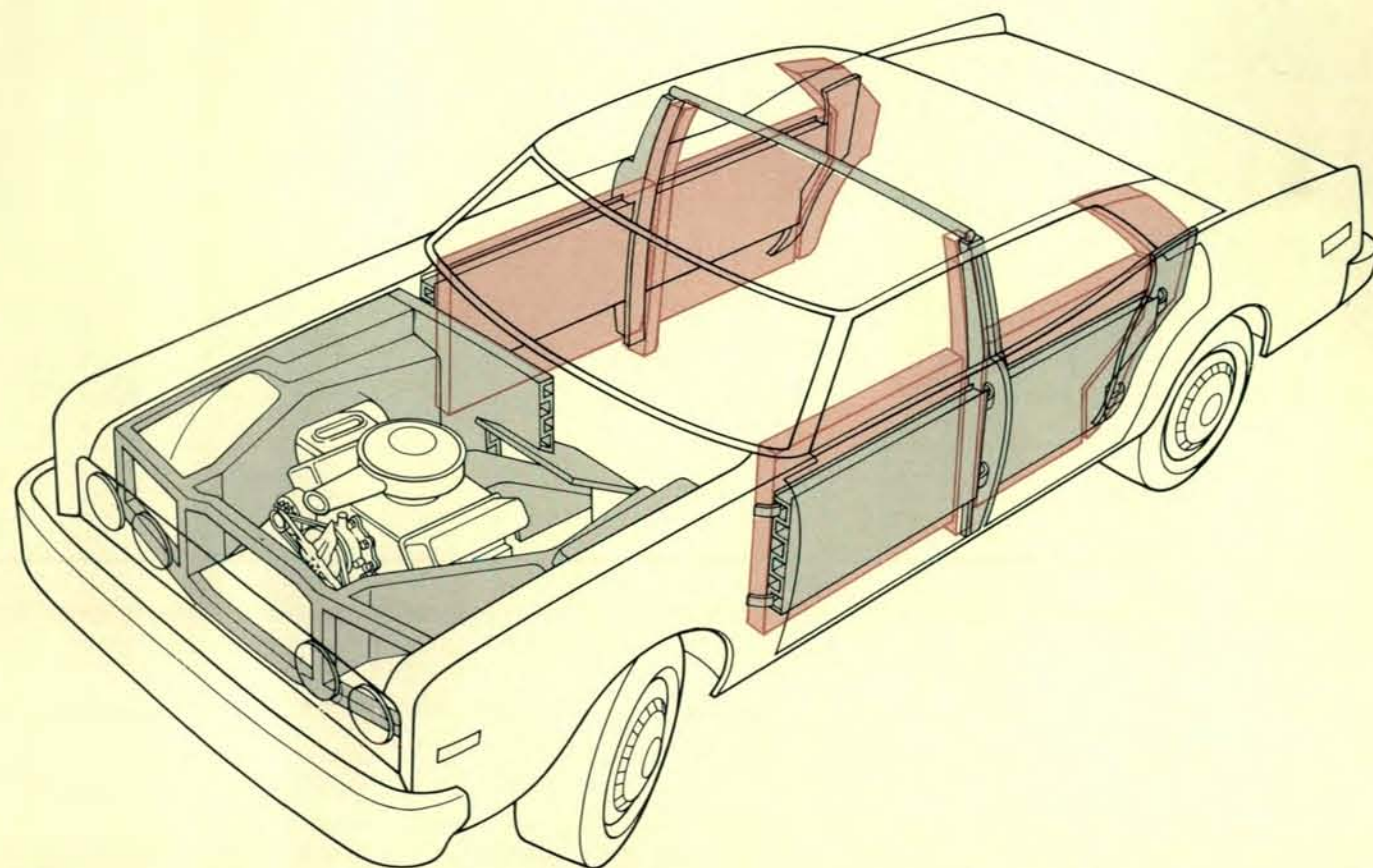
Complessivamente, nell'11,4 % dei 25 000 incidenti considerati il guidatore è stato ferito in modo grave o è morto. Circa la metà di questi incidenti gravi è costituita da collisioni frontali e l'altra metà da collisioni laterali e ribaltamenti. Di conseguenza nel progetto per la sicurezza in caso di urto, la protezione nelle collisioni laterali o nei ribaltamenti richiederà la stessa attenzione che verrà data alla protezione nelle collisioni frontali. Nei casi di collisione laterale e di ribaltamento vi è il rischio particolarmente elevato che gli occupanti vengano sbalzati fuori dalla vettura e quindi va data molta importanza ai dispositivi di ritenuta laterale del passeggero e a quelli atti a prevenire che le porte si trasformino in portelli di espulsione.

Per poter valutare l'efficacia dei dispositivi per la protezione degli occupanti di un'automobile in caso di collisione è stato necessario studiare i limiti di tolleranza del corpo umano rispetto alla decelerazione. Gli esperimenti condotti con i manichini non possono fornire una reale misura delle reazioni umane. Si può però presumere con sicurezza che la gravità degli effetti dipende dal livello e dalla durata della decelerazione. Su questa base la Society of Automotive Engineers ha fissato un « indice di gravità » per quanto riguarda gli effetti relativi alla testa; dalle scarse informazioni desunte dagli incidenti e dagli esperimenti condotti risulta che una esposizione superiore al valore 1000 di questo indice può essere estremamente pericolosa per la vita.

La NHTSA ha fissato certi standard di sicurezza. Nei veicoli sperimentali la decelerazione della testa non deve superare gli 80 *g* (80 volte l'accelerazione di gravità) se dura per piú di 3 millesimi di secondo. Per gli autoveicoli di futura produzione gli standard fissati dalla NHTSA limitano a 1000 l'indice di gravità relativo alla testa, a 60 *g* la decelerazione del torace per tre millesimi di secondo, e a 650 chilogrammi la pressione dinamica che grava sulla parte superiore del femore (anca). Questi criteri si stanno sperimentando su manichini con dimensioni corporee che variano da quelle di una fragile ragazza fino a quelle di un uomo estremamente robusto.

Come si comporta la struttura delle automobili attuali nel proteggere gli occupanti dai valori pericolosi di decelerazione? Questo problema, insieme ad altri, è stato studiato in piú di cento esperimenti di collisione compiuti su automobili dal nostro laboratorio negli ultimi quattro anni. I risultati degli esperimenti indicano che la struttura delle attuali automobili di grosse dimensioni è capace di contenere la decelerazione di un corpo, adeguatamente trattenuto all'interno dell'abitacolo, entro limiti tollerabili in una collisione frontale con un ostacolo fisso a una velocità d'urto compresa tra i 55 e i 65 chilometri l'ora, in una collisione laterale con un albero o un palo a una velocità d'urto fino a circa 16 chilometri l'ora e in una collisione laterale con un autoveicolo di dimensioni analoghe a una velocità d'impatto fino a circa 48 chilometri l'ora. In queste collisioni (tipiche di incidenti che ricorrono nell'ambiente urbano) non c'è stata nessuna grave penetrazione di un veicolo nell'altro, né le porte sono state forzate ad aprirsi.

Perfino a queste velocità moderate però i manichini hanno subito urti con le strutture interne della vettura che sarebbero stati certamente dannosi per le persone. È quindi chiaro che, allo scopo di migliorare nelle automobili attuali la protezione del conducente e dei passeggeri, la priorità dovrebbe spettare all'applicazione di sistemi adatti a limitare lo spostamento degli occupanti dai sedili e ad attutire gli effetti degli urti contro le parti interne, soprattutto quelle sporgenti, e contro i vetri della vettura.

A questo proposito è istruttivo studiare gli esperimenti compiuti con le cinture di sicurezza. Tale analisi è sta-

Le modifiche al telaio apportate dalla Calspan Corporation per migliorare la sicurezza di un'automobile in caso d'urto sono illustrate mettendo a confronto un telaio non modificato (*in alto*) e uno modificato (*in basso*) relativi a un'automobile di dimensioni medie-standard prodotta negli Stati Uniti. Le modifiche consistono nel togliere alcune parti del telaio e nell'aggiungerne altre, tutto allo scopo di aumentare la capacità di assorbimento di energia della parte anteriore e delle fiancate della vettura.

L'automobile modificata costruita sul telaio modificato illustrato nella pagina a fronte comprende variazioni nel frontale, nel supporto del motore, nei pannelli laterali esterni (*in grigio*) e l'aggiunta di parecchi centimetri di spessore di materiale capace di assorbire energia nei pannelli interni del vano passeggeri (*in colore*). Il piano inclinato dietro il motore fa parte di una sistemazione studiata per deviare il motore verso il basso e non dentro il vano passeggeri, in caso di urto frontale.

ta fatta dal nostro laboratorio su un campione di incidenti avvenuti nell'Utah e riguardanti 14 261 passeggeri, 2359 dei quali, al momento della collisione, indossavano cinture di sicurezza. L'indagine ha dimostrato che il mancato impiego della cintura di sicurezza ha aumentato il rischio di morte o di ferite gravi del 100 %. Tra le persone che hanno subito ferite non lievi il numero di coloro che non indossavano la cintura superava del 70 % il numero di chi la indossava e il rischio di incorrere almeno in qualche ferita era del 40 % piú alto tra i non legati che tra i legati. Il vantaggio di indossare la cintura di sicurezza consisteva per tre quarti nella maggior protezione contro le ferite che ci si possono produrre all'interno del veicolo e per l'altro quarto nella ridotta probabilità di essere sbalzati fuori dall'automobile. Lo studio ha indicato che le cinture di sicurezza si sono rivelate utili soprattutto nei casi di collisione laterale.

Le modifiche al frontale comprendono un robusto paraurti in acciaio molto resistente rinforzato da schiuma poliuretanica e una mascherina di alluminio capace di assorbire, in caso di urto, piú energia della mascherina di un'automobile comune.

La paratia tra il motore e il vano passeggeri è stata modificata dai ricercatori della Calspan in modo da ottenere una guida inclinata di spessa lamiera d'acciaio che, in caso d'urto, diriga il motore verso il basso e faccia sí che l'abitacolo salga sul motore.

Con tutta probabilità la combinazione della cintura addominale con quella a bandoliera è la piú efficace. Eppure si deve constatare come la maggior parte degli automobilisti non faccia uso di questi sistemi di sicurezza. Nelle statistiche relative all'Utah nel 1969 solo il 16,5 % delle persone indossava cinture di sicurezza; secondo una stima attuale, nonostante tutte le raccomandazioni fatte dalle autorità, soltanto il 25-35 % usa la cintura addominale e solo il 5 % la combinazione cintura addominale e a bandoliera. Di conseguenza i ricercatori che si occupano del problema della sicurezza dell'automobile si sono rivolti a un'altra alternativa: progettare sistemi di sicurezza « passivi » che non richiedano nessuno sforzo o scelta da parte dell'automobilista.

Questa linea di ricerca si è accentrata soprattutto su un cuscino d'aria che si gonfia immediatamente e automaticamente davanti agli occupanti del veicolo in caso di collisione frontale. La valutazione dell'utilità di tale dispositivo deve tenere in considerazione, tra le altre cose, il tempo disponibile per un suo efficiente funzionamento in caso di incidente. Questo tempo dipende da due fattori: 1) lo spazio che separa il passeggero dalle strutture interne contro cui può urtare durante la collisione dell'autoveicolo e 2) l'entità della deformazione della struttura dell'automobile. Mentre la struttura anteriore si deforma, l'abitacolo continua a spostarsi in avanti nella direzione originale di marcia finché il veicolo è completamente fermo. Lo spostamento in avanti dell'abitacolo durante la collisione, misurato dall'entità del rientramento della struttura esterna, aumenta quindi lo spazio disponibile per l'arresto del movimento del passeggero prima che urti, per esempio, contro il cruscotto.

La struttura a nido d'ape fatta di una pesante lamina di alluminio è la parte che assorbe energia nel pannello aggiunto installato all'interno del vano passeggeri. La struttura di alluminio non è visibile da parte dei passeggeri poiché è ricoperta con pannelli di tipo comune. In alto, un pezzo di materiale a nido d'ape prima dell'urto; in basso, un pezzo dello stesso materiale colpito da una sagoma di acciaio che simula la forma di una parte del corpo umano. Negli esperimenti fatti dalla Calspan Corporation il pannello di alluminio si è dimostrato un po' troppo resistente, cosí che il gruppo di ricerca sta studiando la possibilità di ricorrere a una struttura a nido d'ape di carta.

Supponendo che la decelerazione media di un'automobile in uno scontro sia di 20 *g*, si può calcolare che questo spostamento in un'automobile di grosse dimensioni che collide frontalmente con una barriera piatta a una velocità di 65 chilometri l'ora sia di 80 centimetri. Supponiamo che il passeggero si trovi a 50 centimetri dal cruscotto. La distanza totale disponibile per arrestare il suo movimento sarà allora di 130 centimetri. Con i dispositivi attuali a cuscino d'aria sono necessari almeno 50 millesimi di secondo perché il sistema funzioni: approssimativamente 15 millesimi di secondo per avvertire l'urto e 35 millesimi di secondo per gonfiare il cuscino. Durante questo periodo di tempo il passeggero si sarà spostato in avanti di 89 centimetri (dato che continua a muoversi come un corpo essenzialmente libero alla velocità precedente alla collisione, ossia a 65 chilometri l'ora). Questo riduce lo spazio effettivo disponibile per arrestare il suo spostamento a soli 41 centimetri. In altri termini dei 91 millesimi di secondo necessari a un completo arresto del veicolo in una collisione a 65 chilometri l'ora, restano solo circa 30 millesimi di secondo per l'arresto del movimento del passeggero da parte del cuscino d'aria.

Supponendo che la tenuta esercitata sia uniformemente distribuita per tutto il periodo di decelerazione, per arrestare un oggetto che si muove a 65 chilometri l'ora in uno spazio di 41 centimetri, sono necessari 37 *g* di decelerazione. I cuscini d'aria attualmente sperimentati non funzionano con questi risultati, perché non forniscono una resistenza costante; il cuscino gonfiato esercita solo una resistenza tenue durante i primi 20 centimetri di penetrazione del corpo nel cuscino e poi diventa rapidamente piú rigido, alla stregua di una molla che si indurisce all'aumentare del carico. Per questa ragione è dubbio che i dispositivi a cuscino d'aria attualmente in prova funzionino in maniera accettabile nelle attuali automobili di grosse dimensioni perfino in collisioni che avvengano a 55 o 65 chilometri l'ora. A velocità superiori o in automobili di dimensioni piú piccole il cuscino d'aria certamente non può fornire protezione a meno che il tempo di funzionamento venga considerevolmente ridotto e l'efficacia di ritenuta migliorata.

Ci sono altre ragioni per ritenere che al momento attuale il dispositivo a cuscino d'aria non sembri molto promettente. Il sistema è estremamente complesso e quindi sarebbe piuttosto costoso e non cosí sicuro o di certo funzionamento come si vorrebbe. Nelle versioni attuali offre protezione solo in caso di collisione frontale, non fornendo invece alcun contributo alla

sicurezza in caso di collisioni laterali o di ribaltamenti. Tutto considerato sembra che il concetto del cuscino d'aria, cosí come è stato sviluppato finora, non possa paragonarsi per efficacia alle cinture di sicurezza che si sono invece dimostrate dispositivi di ritenuta particolarmente efficienti.

I dispositivi di ritenuta del passeggero del tipo cosiddetto « attivo » offrono una maggior scelta in fase di progettazione e probabilmente sarebbero in grado di fornire le soluzioni piú soddisfacenti ai problemi della sicurezza dell'automobile, se tutti gli automobilisti potessero in qualche modo essere convinti od obbligati a usarli. Si sono fatti vari tentativi per trasformare un sistema di cinture di sicurezza in un sistema di tipo « passivo » adottando una soluzione che gli automobilisti non potessero rifiutare. Con questo sistema la persona che entra in un'automobile verrebbe automaticamente legata dalla cintura mentre si siede. L'idea però non sembra effettivamente realizzabile perché tutto l'equipaggiamento sarebbe ingombrante e le cinture andrebbero ancorate alle porte che sono un punto non sicuro in alcuni tipi di incidenti.

La ricerca di metodi di protezione dei passeggeri in caso di collisione laterale o di ribaltamento è diretta verso tre obiettivi: l'attutimento degli effetti dell'urto con i fianchi del veicolo (considerato che in tali incidenti questi sono inevitabili), la predisposizione di misure atte a evitare l'espulsione dei passeggeri e il rafforzamento dei montanti delle porte e delle centine per mantenere l'integrità del tetto nei casi di ribaltamento.

Studi recenti compiuti nel nostro laboratorio hanno dimostrato che per un'efficace azione antiurto le sezioni inferiori delle porte dovrebbero essere imbottite con 7-10 centimetri di materiale con alte capacità di assorbire energia. Dovrebbe trattarsi di un materiale piuttosto compatto che ceda alla pressione di penetrazione in modo costante e che riacquisti la sua forma solo lentamente. Sembrano particolarmente adatte allo scopo certe schiume poliuretaniche combinate con strutture di supporto deformabili.

Poiché una spessa imbottitura nelle altre parti delle strutture laterali dell'automobile limiterebbe la visibilità, queste parti dovrebbero essere costruite in modo da non dover sopportare l'intero urto della persona spinta contro di esse (per esempio la testa) ma da scaricarlo o su un finestrino o su un'altra parte assorbente. Si sta studiando un nuovo tipo di vetro per i finestrini laterali e per il lunotto poste-

L'automobile sicura in Italia

Nell'articolo di Patrick M. Miller si è visto come la NHTSA, nell'intento di trovare una soluzione globale al problema della sicurezza dell'automobile, abbia dato inizio al progetto ESV che concentra tutti i suoi sforzi sul tipo di vettura piú diffusa sul mercato degli Stati Uniti, cioè su una berlina 4 porte e 5 posti nella categoria di peso di 4000 libbre (1800 chilogrammi). Considerando tuttavia che il mercato automobilistico americano è attualmente coperto per circa il 15 % da vetture di importazione, con pesi e misure di ingombro sensibilmente minori rispetto alla media delle vetture americane, le autorità statunitensi hanno pensato di estendere il programma ESV anche verso categorie di peso minori e precisamente a tre ulteriori categorie rispettivamente di 3000, 2000 e 1500 libbre (1400, 900 e 700 chilogrammi). Per attuare questo nuovo programma il governo americano ha chiesto la collaborazione di alcune nazioni europee tra le quali vi è l'Italia, e con esse ha sottoscritto accordi bilaterali di cooperazione.

In base agli accordi sottoscritti, all'Italia sono stati assegnati alcuni temi che vengono svolti sotto la direzione e la supervisione del Ministero dei trasporti, dall'Istituto sperimentale auto e motori, dall'Istituto Galileo Ferraris, e dalle nostre due principali case automobilistiche Alfa Romeo e Fiat. L'Alfa Romeo tratta in particolare i due temi del comportamento negli urti e nel ribaltamento, e dei dispositivi di ritenuta per le berline medie con disposizione classica degli organi meccanici (motore anteriore e trazione posteriore), mentre per iniziativa propria prosegue e amplia gli studi sulla sicurezza attiva con particolare attenzione alla ricerca tecnico-sperimentale sulla tenuta di strada dei veicoli. La Fiat dal canto suo collabora alla realizzazione del programma governativo svolgendo i temi che riguardano il comportamento negli urti delle vetture piccole e medie con schema meccanico « tutto avanti » o « tutto dietro », la frenatura, i dispositivi di ritenuta, e la protezione contro il rischio di incendio. Ma l'aspetto piú interessante dei programmi di studio sulla sicurezza intrapresi dalla Fiat riguardano sicuramente le ricerche che essa svolge per proprio conto sulla sicurezza attiva e passiva che l'hanno portata, dopo soli 3 anni di lavoro, alla realizzazione di tre prototipi ESV appartenenti alle categorie di peso di 1500, 2000 e 2500 libbre che corrispondono alle classi di vetture piú diffuse sul mercato europeo. I prototipi discendono infatti da modelli di serie dalle caratteristiche tecniche assai diverse: il piú piccolo dalla « 500 », che ha uno schema meccanico « tutto dietro », il secondo dalla « 128 », con schema meccanico « tutto avanti », e il terzo dalla « 124 », che ha uno schema meccanico tradizionale.

Poiché la vettura sicura non può essere immaginata, neppure in sede teorica, come un mezzo in grado di assicurare una assoluta protezione agli occupanti in qualsiasi tipo di incidente e a qualsiasi livello di severità, il progetto Fiat ha preso le mosse da una accurata analisi degli incidenti stradali reali in modo da individuarne i tipi e i livelli di severità che risultano statisticamente piú frequenti. Le indicazioni tratte da questo studio preliminare non si discostano notevolmente da analoghe indagini effettuate in altri paesi e hanno portato alla conclusione che gli incidenti che provocano il maggior numero di lesioni sono gli scontri frontali contro altri veicoli. La seconda fase del programma è consistita nel sottoporre a prove d'urto alcuni modelli dell'attuale produzione che appartengono alle tre classi di peso prescelte. Le tre macchine sono state sottoposte ad alcune prove d'urto previste dal capitolato americano ESV, con particolare riferimento al problema dello spazio di sopravvivenza in caso di collisione. Oltre che sui modelli in produzione, le prove sono state effettuate anche su modelli modificati nelle strutture in modo da avere indicazioni ed elementi sufficienti per progettare i primi prototipi di vetture sperimentali sicure.

Nel passare alla fase realizzativa dei prototipi, i tecnici si sono prefissi quattro obiettivi che, pur con alcune riserve, sono stati tutti raggiunti: 1. Definizione delle caratteristiche strutturali del corpo vettura ai fini del mantenimento dello spazio di sopravvivenza nei casi tipici di collisione e ribaltamento ad alta velocità. 2. Definizione di un sistema di protezione anteriore e posteriore del corpo vettura negli urti a bassa velocità. 3. Studio del problema dell'aggressività con particolare riguardo ai non protetti (pedoni, ciclisti ecc.) 4. Accertamento dell'incidenza del peso e del costo dei provvedimenti relativi ai tre punti precedenti.

Il primo degli obiettivi è stato risolto in maniera soddisfacente e, allo scopo di mantenere l'indirizzo realistico al quale è improntato tutto il programma, nella realizzazione delle strutture portanti sono stati seguiti gli stessi criteri tecnologici usati nella normale produzione di grande serie.

Per il conseguimento del secondo e del terzo obiettivo, essendo il problema della protezione strettamente legato a quello dell'aggressività, specialmente nei confronti dei non protetti, la soluzione piú logica è stata quella di conferire a un unico sistema entrambi i compiti adottando particolari protezioni elastiche sia per il muso che per la coda del veicolo.

Il prototipo piú piccolo è decisamente il piú interessante e l'unico del quale si disponga di una accurata documentazione (gli altri due sono stati presentati solo nel marzo di quest'anno alla quarta Conferenza sulla sicurezza di Kyoto).

Esso ha ormai un anno di vita (ne sono già stati costruiti 13 esemplari 10 dei quali sono stati sottoposti a prove distruttive) e per molti aspetti merita di essere considerato un po' piú in dettaglio specialmente per quanto concerne il quarto obiettivo del programma Fiat. Esso è infatti il primo esempio di vettura sicura della classe 700 chilogrammi realizzato al mondo e come tale ha permesso di trarre alcune considerazioni e conclusioni di fondamentale importanza per la sicurezza e per il futuro stesso delle piccole cilindrate. Questo specialmente in relazione ai costi di produzione dei piccoli ESV che fin d'ora appaiono proibitivi e destinati a far scomparire dal mercato le vetture appartenenti a questa classe di peso qualora gli standard del capitolato tecnico della NHTSA per gli ESV dovessero essere imposti a tutta la produzione automobilistica europea.

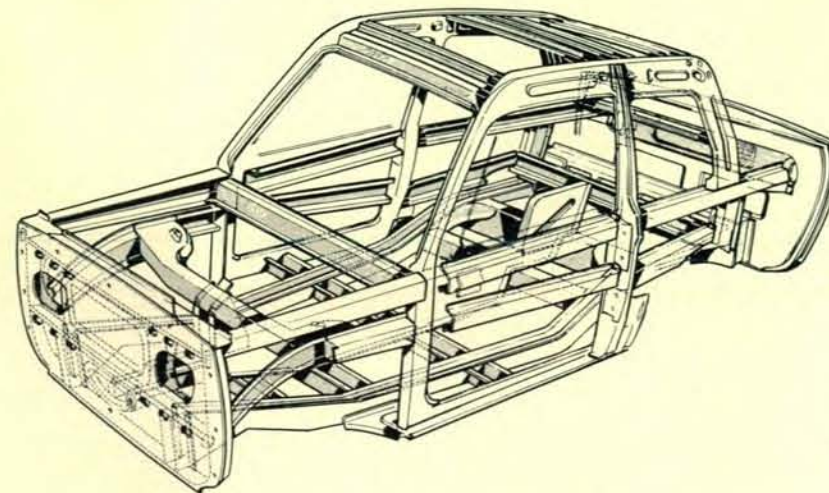
Infatti è proprio sulle vetture di piccola cilindrata che i problemi della sicurezza rischiano di farsi sentire maggiormente costringendo all'adozione di provvedimenti e di dispositivi di costo assolutamente sproporzionati alle possibilità di acquisto dell'attuale vastissima clientela di questo tipo di vetture.

Nella realizzazione del prototipo Fiat che, come abbiamo detto, ha preso le mosse come base dal modello 500 che pesa 550 chilogrammi, si è giunti infatti, pur preoccupandosi solo di soddisfare quella parte delle specifiche ESV relativa allo « spazio di sopravvivenza », a un prototipo che pesa 700 chilogrammi con un aumento di costi e di pesi del 30 % rispetto al modello di partenza. Qualora venissero applicate tutte le specifiche ESV atte a introdurre gli altri indispensabili sub-sistemi di sicurezza, le piccole vetture economiche non avrebbero piú modo di esistere e questo fà sorgere non pochi dubbi e perplessità sull'effettiva ampiezza dei confini entro cui è lecito trattare, al momento attuale, i problemi della sicurezza dell'automobile.

A questo punto è inevitabile porsi una domanda: qual è il valore massimo ragionevolmente accettabile del rapporto costo/beneficio?

A questa domanda non è certo facile dare una risposta definitiva. Entrano infatti in gioco considerazioni ben piú ampie di carattere sociale e politico che implicano inevitabilmente delle scelte di base che è ben difficile quantificare secondo un metro comune. E questo tanto piú in considerazione della realtà attuale in fatto di produzione automobilistica e di circolazione stradale.

Le indagini preliminari sulle statistiche relative agli incidenti stradali hanno rilevato che gli incidenti piú frequenti e piú gravi sono quelli che avvengono per urti frontali e di questo, come si è visto, si è tenuto debito conto nella progettazione degli ESV. Finché si considerano gli urti fra vetture appartenenti alla stessa classe di peso il discorso può essere ancora accettabile; ma che cosa avviene quando si scontrano frontalmente due vetture che pesano rispettivamente 700 e 1800 chilogrammi (evento tutt'altro che infrequente al giorno d'oggi dato che entrambi circolano sulle medesime strade)? Il problema, è chiaro, non può essere risolto proteggendo ulteriormente gli occupanti della piccola cilindrata perché questo sarebbe impossibile anche da un punto di vista tecnologico. L'unica soluzione sarebbe quella di far circolare sulle strade vetture appartenenti alle medesime classi di peso, ma anche questo è attualmente chiaramente impossibile. Siamo dunque ben lungi dal risolvere « globalmente » il problema della sicurezza stradale e questo per motivi obiettivamente insormontabili che affondano le loro radici in un contesto economico e sociale vastissimo.

« Pesi mosca » e « pesi massimi » continueranno ad affrontarsi quindi in impari e cruenti combattimenti sulle nostre strade ancora per molto tempo. *(E.G.)*

Schema della struttura portante del primo prototipo ESV prodotto in Italia dalla Fiat. Questo veicolo appartiene alla classe 1500 libbre e deriva dal modello « 500 » di cui conserva il motore e la disposizione « tutto dietro » degli organi meccanici.

Esperimento col cuscino d'aria eseguito sulla pista sperimentale della Calspan. La serie di fotografie è tratta da un film. La figura sul sedile anteriore è un manichino in parte legato da una cintura di sicurezza addominale. Quando l'automobile urta il palo d'acciaio a destra a una velocità di 66 chilometri l'ora, il manichino viene in un primo tempo spinto in avanti verso il cuscino d'aria e poi scagliato indietro quando il cuscino si gonfia. Esperimenti di questo tipo hanno convinto l'autore che il cuscino d'aria offre una resistenza troppo scarsa nei primi momenti dello scontro e troppo elevata dopo.

riore. Il tipo convenzionale usato attualmente è vetro temperato che in caso di collisione si frantuma in minuscoli frammenti non pericolosi per i passeggeri. Per ottenere finestrini con un alto grado di resistenza sia alla penetrazione sia alla frantumazione, in modo da prevenire l'espulsione dei passeggeri, è allo studio per i finestrini laterali e per il lunotto un tipo di vetro stratificato come quello già adottato per il parabrezza cioè un sottile foglio di acetilbutirraldeide polivinilica, capace di assorbire energia, posto tra due strati di vetro. Nella versione che si sta approntando per i finestrini laterali e per il lunotto il vetro che si trova verso l'interno dell'abitacolo può essere sostituito da una resina poliestere cosí che in caso di collisione non si formino frammenti taglienti.

Abbiamo compiuto ricerche fondamentali sulla struttura esterna dell'automobile progettando e collaudando modifiche intese a far fronte in maniera più efficace ai grossi problemi connessi con la prevenzione di eventuali penetrazioni nello « spazio vitale » dei passeggeri e con la necessità di dissipare l'energia della collisione con il minimo danno per i passeggeri. Ci si è interessati maggiormente delle collisioni frontali e delle collisioni che riguardano i fianchi del veicolo compresi i ribaltamenti.

Sembra che nelle collisioni frontali la struttura convenzionale delle attuali automobili di grosse dimensioni sia adeguata a proteggere i passeggeri quando l'automobile urta una barriera piana fissa a velocità inferiori a 65 chilometri orari, ammesso che i passeggeri siano adeguatamente legati. Abbiamo quindi concentrato i nostri sforzi soprattutto sulle modifiche strutturali progettate per migliorare la protezione in collisioni a velocità comprese tra 65 e 95 chilometri orari. (A queste velocità la deformazione della struttura frontale del veicolo dovrebbe essere compresa tra 91 e 150 centimetri affinché la decelerazione subita dai passeggeri non superi i livelli di tollerabilità.)

Per avere una deformazione strutturale di questa portata si deve fare qualcosa per impedire al motore di essere spinto dentro l'abitacolo. Sono allo studio due soluzioni: l'allungamento del frontale, ovvero della parte sotto il cofano del veicolo, e l'adozione di un piano inclinato che devii il motore, spinto indietro, sotto l'abitacolo.

Nei nostri primi esperimenti di modifiche apportate per rendere un veicolo più sicuro in caso d'urto a velocità di circa 65 chilometri l'ora, abbiamo sistemato un piano inclinato sulla paratia tra il motore e l'abitacolo e abbiamo apportato alla sezione frontale di un'automobile di tipo convenzionale modifiche strutturali atte a rendere la velocità di decelerazione il più possibile uniforme per tutta la sua durata; a questo proposito è stata aumentata la capacità di assorbimento di energia della parte anteriore della struttura e diminuita quella della parte posteriore. Le strutture così modificate sono state poi messe a confronto con quelle di un'automobile di tipo convenzionale in un esperimento d'urto con un palo rigido a circa 95 chilometri l'ora (*si veda l'illustrazione a pag. 25*). Nella vettura modificata il motore, come si sperava, è stato deviato sotto il vano passeggeri invece di venirvi spinto dentro e la decelerazione ha provocato danni relativamente piccoli all'integrità dell'abitacolo. Durante la collisione il massimo di decelerazione dell'abitacolo è stato di 60 *g* nell'automobile modificata contro 85 *g* nell'automobile di tipo convenzionale.

Perfezionamenti sono ora in corso per evitare che questa vettura modificata risulti troppo « offensiva » in caso di collisione con altre automobili. Per questo motivo è stata un po' ridotta la resistenza strutturale del frontale e questo è stato dotato di una mascherina che distribuisce in maniera più uniforme il carico d'urto. Le prove eseguite riguardo a questi perfezionamenti indicano finora che l'automobile sicura non è più offensiva di un'automobile convenzionale che si scontri con una altra di pari dimensioni. Non si sa ancora quali effetti produrrebbe, in caso di scontro, l'automobile modificata su una vettura compatta o subcompatta. È noto come le collisioni tra automobili di grosse dimensioni (1800 chilogrammi) e automobili subcompatte (900 chilogrammi) di tipo tradizionale provochino generalmente gravi danni ai passeggeri di quest'ultima. In relazione al fatto che circa un quarto delle nuove automobili comprate attualmente negli USA sono del tipo subcompatto, i possibili effetti prodotti su tali auto da scontri con veicoli dotati delle modifiche frontali che stiamo ora esaminando dovranno essere studiati attentamente tenendo conto di tutti i tipi di vetture in circolazione prima di poter valutare i benefici di queste modifiche.

Oltre alle modifiche relative alla parte anteriore, stiamo anche collaudando modifiche atte a rinforzare le fiancate dell'automobile. Queste modifiche comprendono traversine tra i longheroni laterali, per rinforzare il telaio sotto il vano passeggeri, crocere poste nelle porte, per sostenere i montanti laterali, e roll bar attraverso il tetto, per impedire che questo ceda in caso di ribaltamento. Per proteggere i passeggeri dagli urti, i punti critici dell'abitacolo sono imbottiti e per i finestrini laterali viene adottato un vetro stratificato ad alto potere di assorbimento di energia (progettato anche per impedire l'espulsione dei passeggeri). L'efficacia di queste modifiche è stata sperimentata in collisioni in cui il veicolo veniva colpito di lato (con un angolo di 90 o di 45 gradi) da una automobile convenzionale di grosse dimensioni che viaggiava a una velocità di 72 chilometri l'ora e anche in collisioni laterali in cui il veicolo modificato colpiva un ostacolo fisso con una velocità d'urto di oltre 30 chilometri orari. In tutti questi casi l'automobile sperimentale si è dimostrata sostanzialmente più efficiente di quelle di tipo convenzionale nel resistere a penetrazioni nel vano passeggeri. Il vetro dei finestrini dell'automobile modificata è rimasto intatto, mentre il vetro non stratificato dell'automobile convenzionale, sottoposto allo stesso urto, si è frantumato.

La frequenza degli incidenti relativa alle diverse parti di un'automobile è illustrata in termini di percentuale di tutti gli incidenti in cui si hanno feriti (*in colore*) e, per ogni gruppo, di percentuale di incidenti considerati gravi (*in nero*). Oltre alle categorie illustrate, gli incidenti in cui l'automobile si ribalta costituiscono il 20,7 % del totale, e di questi il 23,1 % sono gravi. I dati suggeriscono quali componenti strutturali di un'automobile sia opportuno modificare per aumentare la sicurezza in caso d'urto.

Sembra che le modifiche già collaudate apportate alle strutture frontale e laterale non presentino difficoltà per la guida dell'automobile, che potrà venire ancora costruita con i metodi tradizionali. Il programma di ricerca della Calspan Corporation sta procedendo quindi alla costruzione di prototipi che presentino le varie modifiche a carattere protettivo riunite in una automobile di grosse dimensioni di tipo convenzionale. Queste modifiche fanno aumentare il peso dell'automobile di circa il 10 % (approssimativamente 180 chilogrammi).

Gli elementi aggiunti alla vettura per renderla sicura in caso d'urto comprendono schiuma poliuretanica dietro il paraurti anteriore (per assorbire energia in collisioni al di sotto dei 16 chilometri l'ora) e rinforzi che rendano più resistente il telaio in caso di collisioni sia frontali sia laterali (*si vedano le illustrazioni a pag. 26 e a pag. 27*). Per la parte anteriore è stata progettata una mascherina che distribuisca opportunamente le forze d'urto in una collisione frontale con un'altra automobile mentre la parte interna dei parafanghi anteriori, insieme al cofano, contribuisce ad assorbire l'urto. Sulla paratia tra il motore e l'abitacolo è collocato un pannello in acciaio molto resistente che impedisce al motore di penetrare nel vano passeggeri. I montanti, le altre strutture laterali e il tetto dell'abitacolo sono rinforzati. L'interno è imbottito nelle porte, nei montanti e sul padiglione; l'inclinazione del volante è modificata e i finestrini laterali sono di vetro stratificato con materiale plastico.

Queste modifiche alla struttura dell'automobile insieme ad adeguati sistemi di ritenuta dei passeggeri, che andrebbero sempre usati, dovrebbero essere in grado di assicurare agli occupanti un ragionevole grado di protezione nella maggior parte degli incidenti.

I laser a vapori di metallo

Fra i molti sistemi che possono produrre luce laser vi è il metallo vaporizzato miscelato con altri gas. Laser di questo tipo generano radiazione continua a buoni livelli di potenza e in un ampio intervallo di lunghezze d'onda

di William T. Silfvast

Fin dal 1960, quando per la prima volta un laser fu realizzato con successo, l'azione laser è stata ottenuta in molti sistemi fisici differenti. All'originario laser a rubino succedettero altri laser allo stato solido, compresi il laser a neodimio e l'ampia classe dei laser a semiconduttore. Il laser a molecole organiche colorate ha attratto la maggiore attenzione fra i laser a liquido. I laser a gas, che apparirono subito dopo il primo laser a rubino, sono stati estesi a vari tipi speciali quali i laser TEA (TEA significa eccitazione elettrica trasversale a pressione atmosferica), laser chimici e laser gasdinamici. Attualmente vi è un interesse notevole in un'altra classe speciale di laser a gas: i laser a vapori di metallo. In parecchi sistemi a vapori di metallo è stato possibile produrre radiazioni laser continue di molte e differenti lunghezze di onda nelle regioni dello spettro visibile, infrarosso e ultravioletto. Inoltre, i laser a vapori di metallo sono relativamente semplici da costruire e da utilizzare.

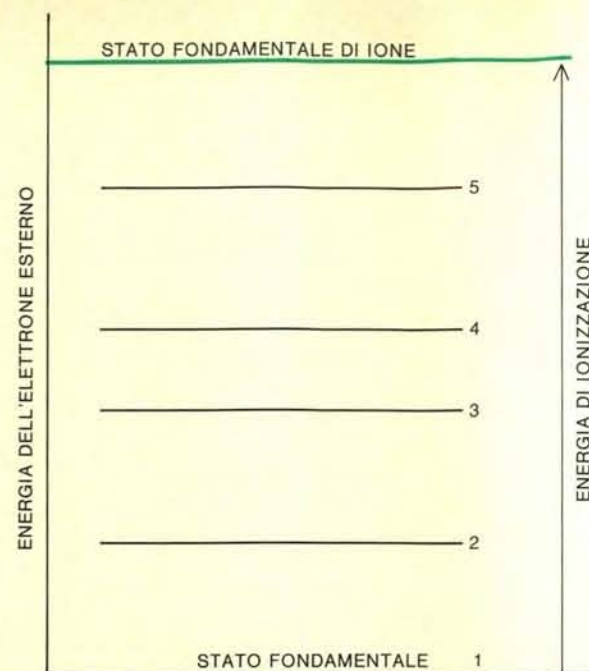
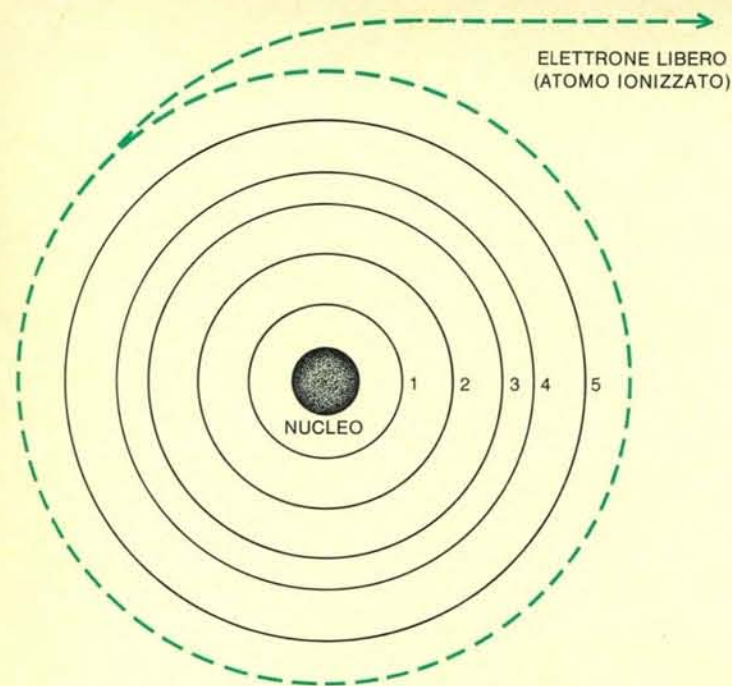
La parola « metallo » suggerisce normalmente l'idea di materiale solido, mentre « vapore » suggerisce l'idea di gas. Perciò il termine « vapore di metallo » può a prima vista sembrare contraddittorio. Il lettore ricorderà, comunque, che ogni elemento chimico può esistere come solido, liquido o gas, in relazione alla temperatura e alla pressione dell'ambiente circostante. Quando il termine « vapore » è usato per descrivere uno stato gassoso, viene inteso che il gas è in presenza dello stato solido o liquido. Un esempio familiare è la coesistenza di vapore d'acqua e di acqua liquida in un recipiente posto su un fornello della cucina. La stessa cosa accade con i laser a vapore di metallo.

Il primo laser di questo tipo è stato quello a vapore di cesio sviluppato da S. Jacobs, P. Rabinowitz e Gordon Gould alla Columbia University nel 1961. Essi fornirono energia al vapore per mezzo di una lampada a flash di elio. Il segnale di uscita del laser era nell'infrarosso. Due anni dopo J. Rigden e A. H. White approntarono il primo laser a vapori di mercurio nei laboratori della Bell Telephone. Gli altri primi lavori inclusero gli sforzi di W. E. Bell e Arnold Bloom della Spectra-Physics, di G. R. Fowles, R. C. Jensen e di me stesso all'Università dell'Utah, di William Walter, Martin Pilthch e dei loro colleghi alla TRG e di A. Tibiloz nell'URSS. Molti di questi lavori erano diretti verso il raggiungimento dell'azione laser attraverso il passaggio di impulsi molto forti di corrente elettrica attraverso un gas rarefatto contenente anche vapori di metallo.

I primi sforzi produssero luce laser pulsata nei vapori di cadmio, zinco, stagno, piombo, manganese, rame, calcio, germanio, indio e oro. I primi laser a vapori di metallo non sostennero favorevolmente il confronto con gli altri tipi di laser nelle applicazioni sperimentali o commerciali poiché gli impulsi laser erano brevi (da 10^{-8} a 10^{-5} secondi) e la loro efficienza era piuttosto bassa. Il laser a elio e neo, che emette radiazione rossa e infrarossa, e il laser ad argo ionizzato, che emette luce blu e verde, attrassero la maggior parte della ricerca e dello sviluppo sui laser a gas poiché entrambi fornivano un segnale d'uscita continuo, in contrapposizione a quello pulsato dei vapori di metallo. Un altro motivo di interesse era che i laser a elio e neo e ad argo ionizzato utilizzavano gas puri, che sono piú facili da maneggiare dei vapori. Inoltre, il laser a elio e neo prometteva un basso costo mentre il laser ad argo ionizzato offriva un segnale d'uscita continuo e ad alta potenza (fino a diversi watt).

In realtà lo sforzo per sviluppare i laser a vapori di metallo non ebbe mai importanza finché Fowles e B. D. Hopkins ottennero un segnale d'uscita quasi continuo di luce laser mediante una scarica a corrente alternata in elio e cadmio. (Il termine « scarica » in questo contesto descrive una corrente elettrica che fluisce in un gas a bassa pressione.) La loro scoperta fu poi successivamente confermata dal nostro lavoro ai laboratori Bell: ottenemmo infatti il primo segnale di luce laser a corrente continua e a onda continua in un vapore di metallo. Le nostre misure indicavano che il laser a elio e cadmio era capace di alta potenza e di notevole efficienza. Gli sforzi per sviluppare i laser a elio e cadmio furono continuati anche da John P. Goldsborough alla Spectra-Physics e da J. R. Fendley e K. G. Hernqvist presso i laboratori della RCA. La scoperta di una nuova riga laser nella regione ultravioletta dello spettro, ove i segnali laser erano stati rari, resero ancor piú interessante il sistema a elio e cadmio.

Come risultato di questi sviluppi, il

Nella fotografia della pagina a fronte sono presentati due laser che funzionano sul principio del vapore di metallo. Quello che emette luce blu è un laser a elio e cadmio e quello che emette luce gialla è un laser a elio e selenio. Una corrente elettrica fornisce energia all'elio. Una parte dell'energia immagazzinata dall'elio è trasferita agli atomi di cadmio o di selenio vaporizzato, che emettono luce quando da livelli di energia elevata cadono a livelli di energia piú bassa, fornendo cosí il segnale d'uscita laser.

Orbite e stati di energia che l'elettrone esterno di un atomo può avere durante il movimento attorno al nucleo. A sinistra vi sono le orbite e a destra i corrispondenti stati di energia. Numeri piú alti indicano energie piú alte. L'energia richiesta per spostare un elettrone su un'orbita o su uno stato di energia piú alti o a distaccarlo dall'atomo (*riga colorata*), ionizzando cosí l'atomo, deve venire da eventi esterni quali la collisione con un elettrone che si muove velocemente. Gli atomi che hanno acquistato energia dopo un tempo piú o meno lungo la riemettono irraggiano i fotoni, o quanti di luce, che un laser amplifica.

concetto di laser a vapori di metallo acquistò un nuovo interesse. Non molto tempo fa, assieme a Marvin B. Klein, ho sviluppato un altro laser a vapori di metallo che crediamo diverrà uno strumento commercialmente importante. Questo laser impiega vapori di selenio e opera con continuità in una scarica a elio e selenio. Esso produce luce laser di molti colori differenti, compreso il blu, verde, giallo, arancio. rosso e infrarosso. Sia il laser a elio e cadmio sia quello a elio e selenio sono relativamente semplici da costruirsi e da utilizzarsi. Attualmente si stanno sviluppando molte applicazioni che sembrano andar bene per le caratteristiche di questi laser.

I lettori ricorderanno che il termine « laser » è l'acronimo della frase inglese « light amplification by stimulated emission of radiation » che significa « amplificazione della luce per emissione stimolata di radiazione ». Nei laser a vapori di metallo e a onda continua che abbiamo studiato, la chiave dell'azione laser è l'immagazzinamento di energia negli atomi di elio. Questa energia ha origine dalla corrente elettrica che viene impiegata per produrre un'opportuna scarica elettrica in un gas consistente di atomi di elio e di atomi di un metallo. Gli elettroni della corrente cedono una parte della loro energia all'elio gassoso, nel quale l'energia è immagazzinata in vari stati eccitati degli atomi di elio. Una parte di questa energia è trasferita successivamente dagli atomi di elio a quelli del metallo durante le collisioni casuali entro la miscela dei gas. Gli atomi di metallo possono essere ionizzati (per perdita di uno o piú elettroni) nei processi di collisione, e molti sono lasciati in stati dai quali possono fornire energia sotto forma di luce laser. Diverrà evidente nel seguito perché i vapori di metallo sono particolarmente utili per questo processo.

Il processo di emissione stimolata di radiazione fu per la prima volta descritto teoricamente da Albert Einstein nel 1917. Si tratta di un processo in cui un fascio di luce di un colore, ovvero di una lunghezza d'onda specifica, può interagire con un atomo in modo tale che l'atomo emetterà luce addizionale che si muove nella stessa direzione e ha la stessa lunghezza d'onda del fascio originario. Se un fascio fosse inserito in un opportuno amplificatore laser, l'amplificatore conserverebbe la direzione e la lunghezza d'onda originaria del fascio entrante mentre l'intensità della luce aumenterebbe. Si può produrre una oscillazione ponendo degli specchi agli estremi dell'amplificatore e costringendo cosí la luce a rimbalzare avanti e indietro fra gli specchi. In un tale oscillatore l'intensità della luce può crescere sino a ordini di grandezza molto piú grandi dell'intensità del segnale di luce originario, producendo un segnale laser altamente direzionale e di colore estremamente puro. La luce viene estratta dal laser attraverso uno degli specchi, che è preparato in modo da trasmettere una piccola percentuale della luce.

La capacità degli atomi di fornire energia dipende dagli elettroni dei singoli atomi. Gli elettroni si comportano come una nuvola di carica negativa attorno a un nucleo carico positivamente e ciascun elettrone occupa uno stato di energia e di momento angolare che non può essere occupato da un altro elettrone; gli elettroni tendono perciò a occupare strati stabili che circondano il nucleo. Gli elettroni dello strato piú esterno sono quelli piú facilmente influenzabili da parte delle forze esterne a causa della loro accessibilità. Questi elettroni esterni possono essere portati a stati di energia piú grande, ma tendono sempre a ritornare al loro stato di energia piú basso: lo stato fondamentale (*si veda l'illustrazione in alto nella pagina a fronte*).

Quando un elettrone è portato a livelli di energia sempre piú grandi, avrà alla fine energia sufficiente per liberarsi dalla forza attrattiva del nucleo e fuggirà dall'atomo; l'elettrone diverrà cosí un elettrone libero con una carica negativa e l'atomo diverrà uno ione con una carica positiva. L'energia necessaria per rimuovere dall'atomo un elettrone (nello stato fondamentale) è chiamata energia di ionizzazione. Ogni specie di atomo ha una sua propria energia di ionizzazione che dipende dal numero di protoni, neutroni ed elettroni che esso possiede. I livelli energetici degli stati degli elettroni degli atomi possono essere misurati in elettronvolt (eV): un elettronvolt è la quantità di energia raggiunta da un elettrone quando è accelerato attraverso una differenza di potenziale di un volt. L'elio ha la piú grande energia di ionizzazione di ogni atomo (24,6 eV) e il cesio la piú piccola (3,9 eV) (*si veda l'illu-*

a

b

c

Emissione e assorbimento di energia da parte di un elettrone. Nell'emissione spontanea (*a*) un elettrone eccitato (*a sinistra*) cade in uno stato di energia inferiore, cedendo energia. L'energia ceduta è emessa sotto forma di un fotone (*in colore*) di determinata energia e perciò di lunghezza d'onda equivalente all'energia ceduta dall'elettrone. Nell'emissione stimolata (*b*), che è il processo che ha luogo in un laser, un fotone avente la differenza esatta in energia fra due livelli energetici di un elettrone collide coll'atomo, inducendo un elettrone del livello superiore a irraggiare la sua energia sotto forma di fotone, cosicché vi sono ora due fotoni invece di uno. L'assorbimento (*c*) ha luogo se l'elettrone è in uno stato inferiore. L'elettrone assorbe il fotone ed è innalzato a uno stato di energia piú alto. Il laser necessita di una larga popolazione ai livelli superiori.

La scarica fornisce un mezzo per eccitare gli elettroni. Se un gas quale l'elio è messo in un recipiente di vetro a bassa pressione e un'alta tensione è inserita fra i due elettrodi, viene prodotto un campo elettrico (*in colore*), che applica una forza alle cariche e induce un flusso di corrente fra gli elettrodi. Gli elettroni, avendo una massa molto piccola, sono accelerati a velocità estremamente alte. In un laser a vapori di metallo essi collidono con gli atomi di elio, innalzando gli elettroni delle orbite esterne di questi atomi a uno stato di energia piú alto. Questa energia è quindi trasferita attraverso collisioni agli atomi neutri (*circoletti*), fornendo cosí le basi per l'azione laser.

L'area luminosa fra gli elettrodi è la regione della scarica a gas di un laser a elio e cadmio realizzato nel laboratorio dell'autore. L'anodo è a sinistra, il catodo a destra. Il cadmio che deve essere vaporizzato per fornire gli atomi di vapore di metallo al sistema è l'oggetto sottile e scuro entro il tubo, a breve distanza dalla destra dell'anodo. Il fascio di luce appare alla destra e alla sinistra delle finestre inclinate alle estremità del tubo. Questo piccolo laser (la distanza fra le finestre è soltanto 20 centimetri), produce un segnale d'uscita continuo di due milliwatt di luce blu a una lunghezza d'onda di 4416 angstrom. La luce che viene emessa è estremamente pura e di alta intensità.

strazione in questa pagina).

Quando un elettrone è in uno degli stati di energia piú alti, cercherà sempre di raggiungere uno degli stati di energia piú bassi, a volte saltando direttamente allo stato fondamentale, a volte scendendo da un livello al successivo come lungo una scalinata. In tali transizioni l'energia deve essere conservata, e quindi l'atomo emette un fotone, cioè un quanto di luce, che ha energia corrispondente esattamente alla differenza fra le energie dei due stati dell'elettrone. La frequenza, ovvero la lunghezza d'onda della luce emessa, è correlata alla differenza di energia. Un fotone rosso ha una energia di circa 2 eV, un fotone blu una energia di circa 3 eV.

Gli elettroni di certi livelli decadono (ossia cadono a uno stato piú basso) piú facilmente degli elettroni di altri livelli. Ciascuno stato elettronico eccitato dell'atomo ha una vita media caratteristica che indica il tempo medio impiegato da un elettrone per cadere a un livello piú basso e per emettere quindi un fotone. Gli stati piú eccitati hanno vita media di circa 10^{-8} secondi.

In tutti gli atomi vi sono alcuni stati o livelli eccitati da cui un elettrone non può decadere facilmente attraverso la cessione di un fotone. Tali atomi devono perciò aspettare altri mezzi per cedere la loro energia, per esempio le collisioni con altri atomi o con le pareti del sistema. Gli elettroni in stati energetici di questo tipo tendono a rimanervi per periodi di tempo relativamente lunghi (0,001 secondi o piú) e in tal caso si dice che gli elettroni sono in stati metastabili. Questi stati hanno un ruolo importante nell'immagazzinare l'energia che può essere impiegata durante il processo di eccitazione di un laser a vapori di metallo.

Il normale decadimento radiativo da uno stato elettronico piú alto a uno piú basso viene chiamato emissione spontanea. Esistono anche processi che possono costringere un elettrone atomico a passare in uno stato piú alto oppure a saltare in uno stato piú basso. Un esempio di costrizione è fornito quando un fotone collide con un atomo ed eccita l'elettrone piú esterno a un livello piú alto; ciò può accadere quando l'energia, ovvero la lunghezza d'onda del fotone incidente, corrisponde esattamente alla differenza in energia fra lo stato in cui si trova l'elettrone e qualche possibile stato piú alto. Questo processo è detto assorbimento poiché il fotone è realmente assorbito dall'atomo e tutta l'energia del fotone viene impiegata per far passare l'elettrone in uno stato piú alto.

Analogamente, l'elettrone stimolato può muoversi verso un livello piú basso, a patto che un tale livello esista e che la differenza tra i due livelli corrisponda esattamente all'energia del fotone. Questo è il processo di emissione stimolata. La energia ceduta dall'elettrone nel saltare a uno stato piú basso viene utilizzata nella creazione di un fotone addizionale con le stesse caratteristiche di quello collidente (*si veda l'illustrazione in basso a pagina 36*).

Un gran numero di atomi può produrre un aumento di luce entro un amplificatore laser se la popolazione di elettroni negli stati eccitati degli atomi è preparata in modo opportuno. Si considerino due livelli eccitati di un sistema di atomi identici aventi gli elettroni divisi fra il livello superiore e quello inferiore. Se un fascio di luce di lunghezza d'onda corrispondente alla differenza in energia fra i due livelli è costretto a passare attraverso il mezzo, verrà amplificato qualora vi siano piú atomi con elettroni nello stato superiore, mentre verrà assorbito qualora vi siano piú atomi con elettroni nello stato inferiore. La condizione di avere piú atomi nello stato superiore è chiamata « inversione di popolazione » (poiché essa va contro il normale processo naturale, che tende a tenere piú elettroni a energie piú basse piuttosto che a energie piú alte). La ricerca di nuovi sistemi laser non è quindi facile, poiché bisogna trovare un sistema che contrasti la tendenza naturale ad avere piú atomi con elettroni nel livello superiore piuttosto che atomi con elettroni nel livello inferiore.

Uno dei modi migliori per fornire energia a un gas allo scopo di ottenere una inversione di popolazione consiste nell'utilizzare una scarica elettrica (*si veda l'illustrazione in alto a pagina 37*). Una scarica inizia quando fra due elettrodi posti entro un recipiente di vetro riempito con un gas a bassa pressione si stabilisce un'alta tensione. La tensione produce un campo elettrico nel gas, e il campo genera un flusso di corrente elettrica fra gli elettrodi in quanto sottopone le cariche elettriche a una forza. Gli elettroni carichi negativamente sono accelerati verso l'anodo, mentre gli ioni carichi positivamente sono accelerati verso il catodo. Gli elettroni, che sono molto piú leggeri degli ioni, sono accelerati a velocità altissime (10^8 centimetri al secondo) e perdono parte della loro energia collidendo con gli atomi del gas. Alcuni degli elettroni acquistano energia sufficiente per ionizzare altri atomi creando cosí molti elettroni liberi che aiutano a sostenere la scarica.

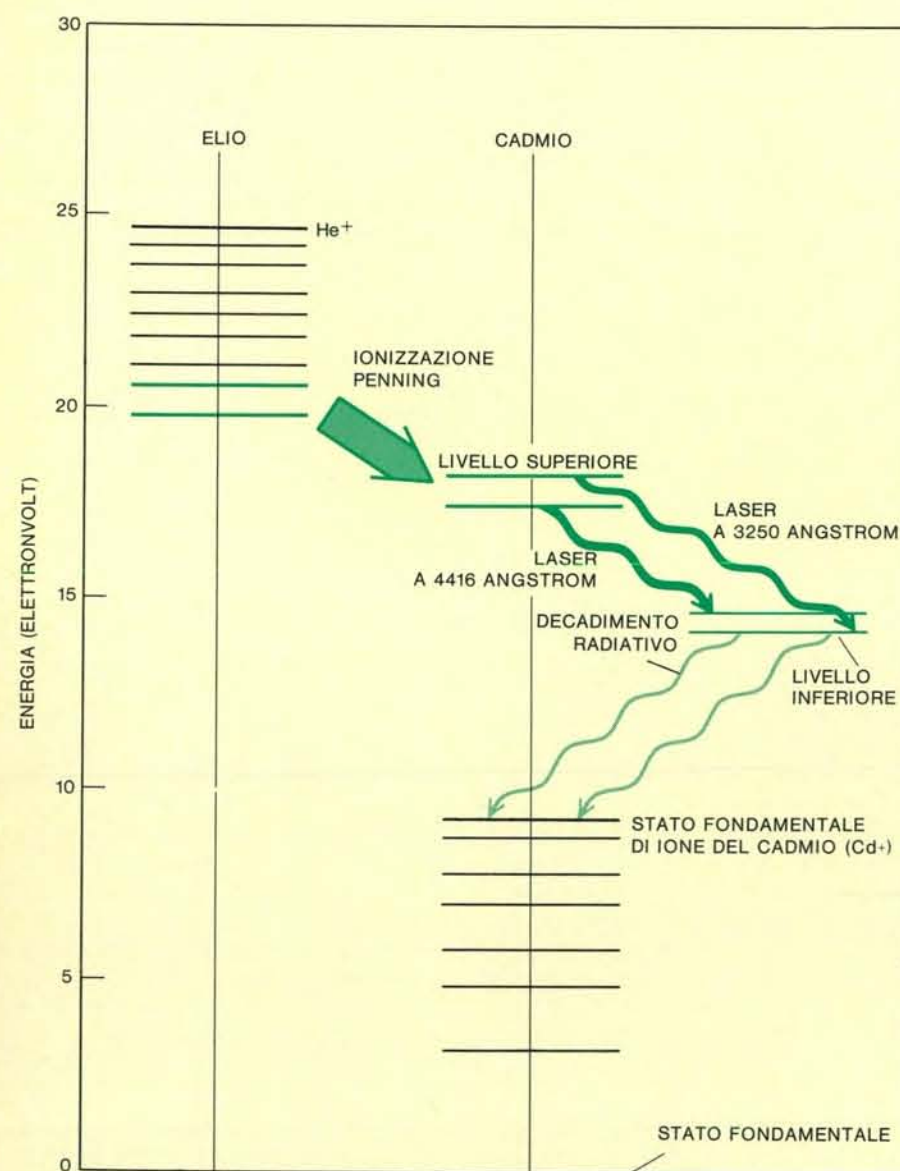
Vi sono due tipi di collisioni fra gli elettroni e gli atomi. Uno è il tipo elastico, che assomiglia all'impatto di due palle da biliardo: il leggero elettrone colpisce il piú pesante atomo, dandogli una lieve spinta. In questo tipo di collisione l'elettrone perde solo una piccola frazione della sua energia a causa della sua piccola massa. Nell'altro tipo di collisione, detto anelastico, l'elettrone coinvolge uno degli elettroni dello strato esterno dell'atomo portandolo a uno stato piú alto. In questo caso l'elettrone libero può perdere una larga frazione della sua energia perché molti stati atomici sono separati da energie di diversi elettronvolt.

L'elio, fra tutti gli atomi, ha il piú grande divario di energia fra lo stato fondamentale e il primo stato eccitato (*si veda l'illustrazione nella pagina a fronte*). Pertanto, quando una corrente elettrica viene immessa nell'elio per produrre una scarica, gli elettroni liberi devono essere accelerati a una energia superiore a 20 eV prima che essi possano prender parte a collisioni anelastiche con gli atomi di elio. All'estremo opposto, invece, gli elettroni liberi di una scarica nel cesio puro necessitano di una energia poco superiore a un elettronvolt per prendere parte alle collisioni anelastiche con gli atomi di cesio. Pertanto la ben distinta e diversa struttura dei livelli energetici dell'elio e del cesio aiuta a produrre un'alta energia media degli elettroni nella scarica a elio e una bassa energia media nella scarica a cesio.

L'alta energia degli elettroni nella scarica a elio induce molti atomi di elio a essere eccitati a stati molto energetici. Tre di questi stati sono metastabili, ossia hanno lunga vita media. Essi sono chiamati rispettivamente stato metastabile di singoletto, stato metastabile di tripletto e stato fondamentale di ione. Gli stati metastabili di singoletto e di tripletto hanno vita media lunga poiché non hanno mezzi per irradiare la loro energia. Lo stato fondamentale di ione, che è il piú basso stato di

Stati energetici possibili per l'elettrone esterno di un atomo di elio e di un atomo di cesio. A causa della quantità di energia richiesta per innalzare l'elio al primo stato eccitato, che è molto piú alto del primo stato eccitato del cesio, l'elio è utile per accumulare energia in quantità notevoli. Inoltre, diversi stati (*in colore*) sono metastabili, ossia hanno vita media relativamente lunga durante la quale possono trattenere l'energia.

I livelli di immagazzinamento dell'elio, cioè i livelli metastabili capaci di rilasciare agli altri atomi l'energia immagazzinata, a confronto con i livelli di energia di diversi altri sistemi atomici. A causa di restrizioni nel processo di trasferimento di energia durante la collisione, soltanto i livelli di ione degli altri atomi possono ricevere energia dai livelli metastabili dell'elio. L'argo e il fluoro non hanno raggruppamenti di livelli di ione che siano sotto i livelli metastabili dell'elio e che abbiano altre proprietà richieste per l'azione laser. Lo zinco e il piombo hanno invece molti livelli di ione sotto gli stati metastabili dell'elio.

Il trasferimento di energia da un atomo metastabile di elio (*colore scuro*) a un atomo di cadmio in un laser a elio e cadmio richiede una collisione. A sinistra gli atomi sono mostrati prima della collisione a destra dopo di essa. Nel trasferimento di energia l'atomo di elio ritorna al suo stato fondamentale e l'atomo di cadmio è ionizzato. La differenza di energia fra lo stato metastabile dell'elio e lo stato di ionizzazione del cadmio è portata via da un elettrone. Il processo è noto come ionizzazione Penning.

La ionizzazione Penning in un laser a elio e cadmio eccita gli atomi di cadmio a stati ionizzati. La radiazione dai livelli superiori degli ioni di cadmio inizia l'azione laser blu a 4416 angstrom e quella ultravioletta a 3250 angstrom. Gli atomi decadono dai livelli piú bassi piú velocemente che da quelli piú alti, cosicché è mantenuta la richiesta inversione di popolazione degli ioni eccitati. Esistono anche livelli sopra i livelli laser.

energia dell'elio ionizzato (elio con un elettrone rimosso dall'atomo), ha una lunga vita media perché deve catturare un elettrone prima di poter irradiare la sua energia. Molti atomi di elio eccitati tendono a finire in uno dei tre stati metastabili, creando nella scarica un grande serbatoio di energia elettronica.

Se altri atomi, che possono essere considerati come atomi di impurità nel caso che la loro concentrazione sia relativamente bassa, sono introdotti in una scarica a elio, allora le collisioni casuali nella scarica misceleranno gli atomi di impurità con gli atomi di elio, e i due tipi di atomi collideranno piuttosto frequentemente. Durante tali collisioni l'energia immagazzinata negli atomi metastabili di elio può essere trasferita agli atomi di impurità, lasciando gli atomi di impurità in uno stato elettronico eccitato da cui possono successivamente emettere radiazione quando precipitano ai livelli piú bassi. Purché certe condizioni siano soddisfatte, è possibile ottenere una inversione di popolazione fra i due livelli eccitati degli atomi di impurità.

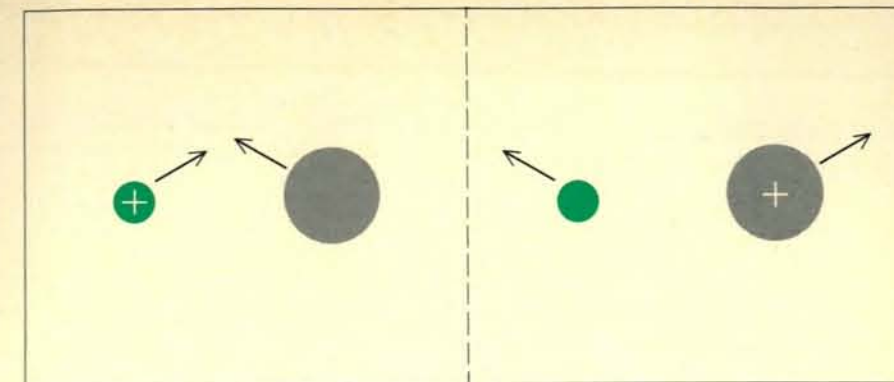
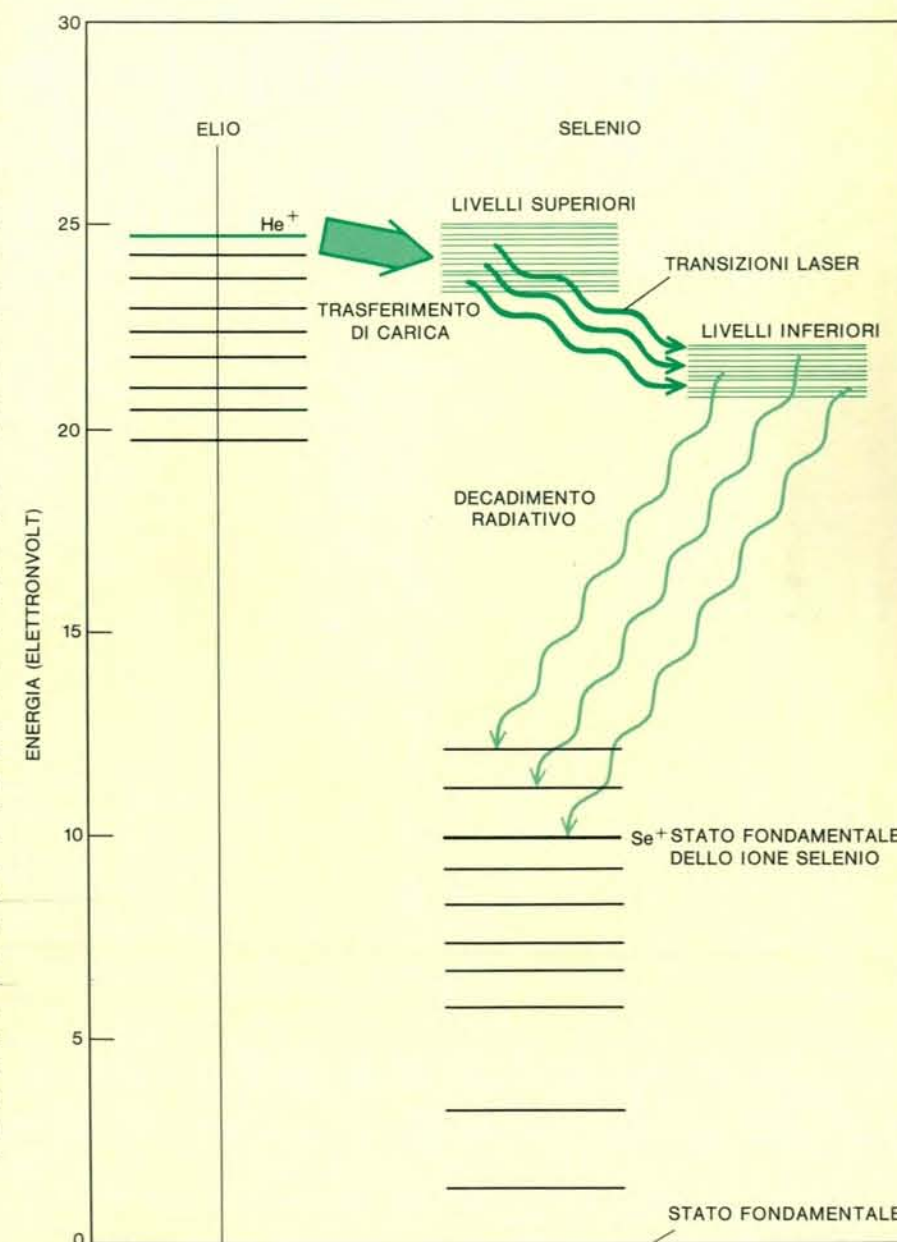
Una condizione è che l'atomo di impurità deve avere livelli opportuni di energia di ione. Per essere adatti devono avere un'energia uguale o minore dell'energia dei livelli metastabili dell'elio, poiché gli atomi di elio metastabili non possono fornire un'energia maggiore di quella che hanno inizialmente. A causa delle restrizioni che governano il processo di scambio di energia durante la collisione solo i livelli di ione di altri atomi possono ricevere energia dai livelli metastabili dell'elio. Il neo è la sola eccezione, e la ragione di ciò sta nel fatto che esso ha un alto potenziale di ionizzazione. (I laser a elio e neo coinvolgono uno scambio di energia dai livelli metastabili dell'elio ai livelli dell'elio neutro.)

Un secondo fattore riguarda la vita media dei livelli di energia di ione relativa agli atomi di impurità coinvolti nell'inversione di popolazione. Il livello superiore del sistema a due livelli con una inversione di popolazione irraggia energia corrispondente esattamente alla differenza fra i due livelli. Gli elettroni che hanno cosí reso energia per irradiazione finiscono al livello piú basso. Se essi rimanessero in tale livello, sorgerebbe subito una situazione nella quale vi sarebbero piú atomi nel livello inferiore che in quello superiore, e vi sarebbe un assorbimento invece di un guadagno. Quindi, per avere il funzionamento laser continuo, bisogna trovare atomi di impurità con combinazioni di livelli tali che quello inferiore abbia una vita media breve e quello superiore sia popolato preferenzialmente. Questa necessità scarta lo stato fondamentale di ione quale possibile livello laser, poiché tutti gli stati fondamentali di ione sono metastabili. La coppia dei livelli laser di ione deve avere energia superiore allo stato fondamentale di ione ma inferiore agli stati metastabili di elio da cui ricevono energia.

Una ultima condizione è che gli atomi di impurità siano vaporizzati. In quanto vapore essi sono in uno stato gassoso, cosicché i loro livelli di energia corrispondono ai livelli discreti degli atomi o delle molecole singole in contrasto con i livelli allargati di energia di un liquido o di un solido. Alcuni elementi, che richiedono una temperatura di 2000 °C o piú per essere vaporizzati in maniera significativa, non sono di uso pratico per i laser a vapori.

Con queste condizioni in mente si può esaminare la tavola periodica per vedere quali elementi dovrebbero divenire buoni candidati per i laser. Risulta che gli atomi di un certo numero di metalli soddisfano le richieste. Le possibilità possono essere illustrate attraverso una comparazione degli stati di energia dell'elio con quelli di due elementi non metallici (argo e fluoro) e di due metalli (piombo e zinco).

Nel caso dell'argo soltanto lo stato fondamentale di ione può essere raggiunto dai livelli metastabili dell'elio, e ciò elimina chiaramente l'argo quale candidato opportuno per questo genere di laser (*si veda l'illustrazione a, pagina 39*). Il fluoro ha due livelli sopra lo stato fondamentale di ione che possono ricevere energia dai livelli metastabili di elio, ma anch'essi sono livelli metastabili e quindi non forniscono una base per un efficiente irraggiamento di energia. Piombo e zinco, che sono tipici atomi di metallo, hanno basse energie di ionizzazione e anche un numero considerevole di livelli fra la energia di ionizzazione e i livelli metastabili dell'elio che rendono possibile l'azione laser. Naturalmente i livelli di energia di un metallo, per essere completamente soddisfacenti, devono anche avere vite medie opportune e buone velocità di trasferimento allo scopo di ottenere un efficiente trasferimento di energia dall'elio.

Guardiamo ora con maggior attenzione ai due laser a vapori di metallo che sono stati descritti brevemente all'inizio di questo articolo: il laser a elio e cadmio e quello a elio e selenio. Il

La ionizzazione per trasferimento di carica in un laser a elio e selenio coinvolge una collisione fra un atomo di elio ionizzato (*in colore*) e un atomo di selenio neutro (*in grigio*). Un elettrone passa dal selenio all'elio, lasciando l'elio neutro e il selenio ionizzato.

L'azione laser in una scarica a elio e selenio si ha quando un atomo di selenio riceve energia da uno ione di elio attraverso la cessione di uno dei suoi elettroni e innalzando un altro degli elettroni dell'orbita esterna a uno dei 13 livelli laser superiori. Da qui l'energia è irradiata a uno degli 11 livelli inferiori, dando inizio all'azione laser.

a

ANODO

CATODO

ALIMENTATORE DEL RISCALDAMENTO

SORGENTE DEL METALLO

METALLO CONDENSATO

b

ANODO

CATODO

CAMMINO DI RITORNO

SORGENTE DEL METALLO

c

ANODO

CATODO

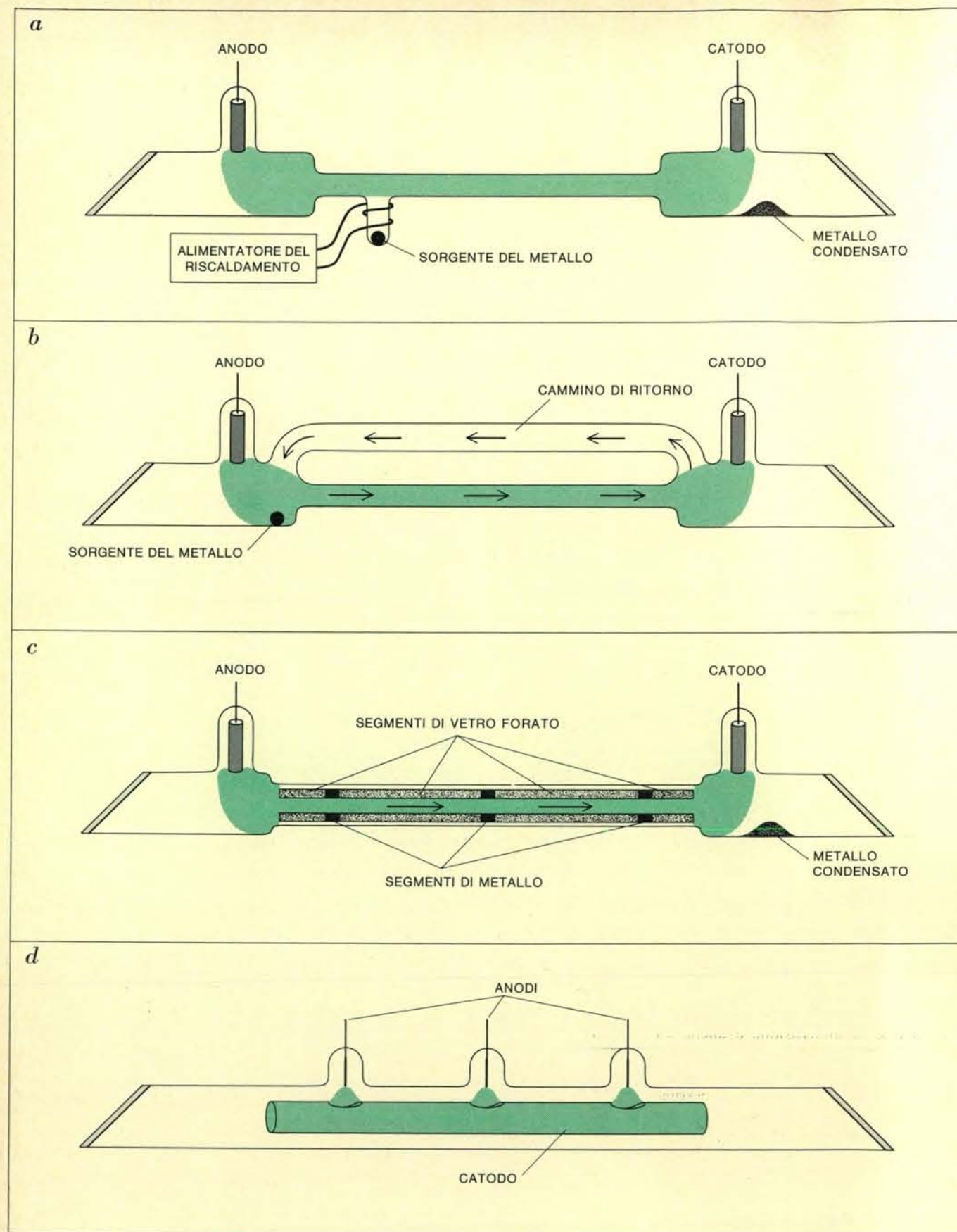
SEGMENTI DI VETRO FORATO

SEGMENTI DI METALLO

METALLO CONDENSATO

d

ANODI

CATODO

Tubi laser progettati per funzionare con continuità quali laser a vapore di metallo. In ciascuno dei casi l'area colorata indica la regione della scarica. Un tubo (*a*) impiega un processo di flusso chiamato pompaggio cataforesico. Il campo elettrico della scarica distribuisce il vapore di metallo. Un tubo di ritorno (*b*) ricicla il vapore continuamente. Il disegno semplificato omette il riscaldatore, le scariche necessarie per impedire al metallo di condensarsi e la griglia che blocca la scarica dal cammino di ritorno. Nel tubo a foro segmentato (*c*) segmenti di metallo sono interposti fra segmenti di vetro e il calore della corrente della scarica vaporizza il metallo. Il tubo a catodo cavo (*d*) ha un lungo catodo che riscalda il metallo.

laser a elio e cadmio utilizza i livelli metastabili di tripletto e di singoletto dell'elio (*si veda l'illustrazione in basso a pagina 40*). Nel processo di trasferimento dell'energia gli atomi metastabili di elio collidono con gli atomi neutri di cadmio che sono nello stato fondamentale, e l'energia metastabile dell'elio è ceduta al cadmio. Al fine di conservare l'energia e la quantità di moto il cadmio si ionizza, fornendo così un elettrone libero alla scarica. Con le tre particelle separate, l'unica maniera in cui un elettrone può realizzare l'equilibrio dell'energia e della quantità di moto consiste nel portare una energia che sia la differenza esatta fra il livello di energia del cadmio e il livello metastabile dell'atomo di elio. Questo tipo di collisione e di ionizzazione è detto ionizzazione Penning in onore del compianto F. M. Penning che scoprí per primo il processo.

Il sistema a cadmio soddisfa facilmente le richieste di vita media che sono state discusse. I livelli laser superiori hanno vite medie di circa 0,5 microsecondi, mentre le vite medie ai livelli inferiori sono solo di pochi nanosecondi, cioè circa 230 volte piú brevi. Inoltre, da misure fatte sulle velocità di eccitazione ai livelli superiori e inferiori, sembra che il trasferimento di energia dall'elio al livello superiore proceda con una velocità di circa tre volte piú grande della velocità di trasferimento ai livelli inferiori. Un altro interessante comportamento del sistema è il fatto che il cadmio richiede di essere riscaldato a soli circa 250 °C per fornire una densità di vapori di cadmio sufficiente per il trasferimento di energia, densità utile per procedere a un ritmo abbastanza grande nell'azione laser. A causa del processo di eccitazione relativamente efficiente, e anche delle vite medie favorevoli dei livelli laser superiore e inferiore, il laser a elio e cadmio ha prodotto 300 milliwatt di potenza laser a una lunghezza d'onda blu di 4416 angstrom e 50 milliwatt nell'ultravioletto a 3250 angstrom. Altri laser continui che usano collisioni Penning per produrre il trasferimento dell'energia da parte dell'elio sono stati sperimentati con zinco, stagno e piombo.

Il laser a elio e selenio deriva la sua energia dallo stato fondamentale degli ioni metastabili di elio. Gli ioni di elio collidono con gli atomi neutri di selenio che sono stati scissi dalla loro forma di vapore molecolare attraverso collisioni con gli elettroni presenti nella scarica, e poiché gli ioni di elio hanno un elettrone mancante essi possono prontamente assorbire un elettro-

ne dal selenio. L'energia è trasferita agli atomi di selenio, lasciandoli in uno stato ionizzato, e gli ioni di elio divengono atomi di elio neutri nello stato fondamentale.

Questo tipo di reazione è nota come processo di trasferimento di carica o di scambio di carica poiché la carica elettronica è trasferita dal selenio all'elio. Siccome nel processo di collisione non viene liberato un elettrone, a differenza di quanto avviene nella ionizzazione Penning, i soli livelli del selenio che possono accettare energia in modo efficiente sono quelli vicini all'energia degli ioni di elio. Il selenio ha 13 livelli laser che hanno energia sufficientemente vicina allo stato fondamentale dello ione di elio e che sono utili per ricevere energia e servire quali livelli laser superiori (*si veda l'illustrazione in basso a pagina 41*). Questi livelli irraggiano luce e decadono verso 11 stati inferiori che servono quali livelli laser inferiori. I livelli laser inferiori hanno tutti una rapida rotta di fuga verso lo stato fondamentale di ione in modo da fornire il necessario canale verso una inversione di popolazione.

Il laser a elio e selenio, con le molte lunghezze d'onda e con la relativamente alta efficienza, ha prodotto una potenza combinata fino a 250 milliwatt su sei lunghezze d'onda di luce blu e verde; le potenze alle singole lunghezze d'onda sono state di 50 milliwatt. Inoltre sono state simultaneamente ottenute da un singolo laser 19 lunghezze d'onda, ricoprenti gran parte dello spettro visibile. Altri vapori di metallo che hanno prodotto un segnale di uscita laser continuo per mezzo del processo di trasferimento di carica includono il cadmio, lo zinco e il tellurio.

Poiché i vapori di metallo esistono in congiunzione sia con lo stato solido sia con quello liquido del metallo, per contenere il metallo solido o liquido e per distribuire il vapore entro la scarica sono richiesti speciali tubi laser. Inoltre è di solito necessario uno strumento per riscaldare il metallo alla appropriata tensione di vapore. Per soddisfare queste richieste sono stati progettati diversi tipi di tubi per la scarica a gas correntemente ne vengono usati quattro tipi (*si veda l'illustrazione a pagina 42*).

Un tipo di tubo contiene una singola sorgente di vapori di metallo posta vicino al terminale della tensione positiva (l'anodo). Attraverso un riscaldatore separato la sorgente è riscaldata a circa 250 °C per il cadmio e selenio onde vaporizzare il metallo nella regione di scarica. Quando il vapore raggiunge la scarica, alcuni degli atomi sono ionizzati e gli ioni positivi di metallo sono attratti verso il catodo da parte del campo elettrico. Il risultato complessivo è un flusso di vapori di metallo verso il catodo.

La scarica fornisce calore sufficiente a impedire al vapore di metallo di condensare nella regione del foro del tubo, per cui il metallo si condensa quando raggiunge la regione del catodo dove non vi è scarica e la temperatura è bassa. Perciò vi è un continuo flusso di metallo attraverso il foro: la regione dove si ottiene l'amplificazione laser nella quale la corrente elettrica fornisce la necessaria inversione di popolazione. Questo processo di flusso è chiamato pompaggio cataforesico. La sua applicazione ai laser a vapori di metallo fu sviluppata indipendentemente dai gruppi della Spectra-Physics e della RCA e da Thomas Sosnowski dei laboratori Bell. Nel laser a elio e cadmio il processo consuma cadmio a una velocità di circa un grammo per 700 ore di funzionamento del laser.

Il secondo tipo di sistema usa il metallo piú e piú volte. Il vapore fluisce dall'anodo al catodo nella scarica del laser e quindi si diffonde indietro verso l'anodo attraverso un cammino di ritorno. Questo laser fu sviluppato da Hernqvist alla RCA per trarre profitto di speciali isotopi di cadmio che danno un guadagno piú alto di quello che dà la distribuzione isotopica naturale. Gli isotopi speciali sono cari, ma in questo tubo ne è richiesta solo una piccola quantità poiché il metallo viene riutilizzato.

Un altro approccio, che ho sviluppato assieme a Leo Szeto ai laboratori Bell, impiega un catodo con un foro segmentato. Una o piú pallottole cilindriche di metallo percorse da un foro che si combina con la regione del foro entro il tubo, sono intervallate da segmenti di vetro forati nella regione di scarica. La scarica elettrica non solo fornisce gli elettroni ad alta energia per eccitare l'elio ma riscalda anche il metallo alla giusta tensione di vapore. Il processo di cataforesi distribuisce quindi il metallo entro i segmenti di vetro forati. Il dispositivo, che è poco costoso, è stato sperimentato con successo usando cadmio, selenio e zinco.

Il quarto tipo di laser – il laser a catodo cavo – ha un catodo cilindrico entro cui il metallo è riscaldato. La corrente fluisce da anodi multipli verso il catodo e passa attraverso una fessura del catodo per produrre una incandescenza uniforme nell'area del catodo. Questi tipo di laser è piú adatto per il tipo di eccitazione a trasferimento di carica che al tipo Penning. Esso non ha però ancora raggiunto i livelli di potenza che possono essere prodotti dagli altri tre sistemi. Il primo segnale d'uscita continuo da una struttura a catodo cavo fu ottenuto indipendentemente da diversi gruppi in Giappone, in URSS e in USA.

Per un laser piccolo e non costoso che operi nella regione blu o ultravioletta dello spettro vi sono molte applicazioni potenziali. Per esempio, parecchie molecole sono molto sensibili alla radiazione in questa regione e attraverso il cambiamento degli stati energetici delle molecole per effetto della luce laser si può valutare la struttura chimica della sostanza. Analogamente, molti fotorivelatori raggiungono la risposta massima in luce blu, e possono quindi essere usati con i laser a luce blu in applicazioni quali esaminare una pagina di testo, trasformare le diverse intensità di bianco e nero in segnali elettrici e trasmetterli a terminali distanti, dove il materiale può essere riprodotto.

Una applicazione attualmente in studio ai laboratori Bell da parte di Fritz Froehlich e dei suoi colleghi usa un piccolo laser ultravioletto a elio e cadmio simile al dispositivo a foro segmentato che ho descritto. Il sistema, chiamato lavagna lontana, dovrebbe permettere la trasmissione di informazioni scritte utilizzando una normale linea telefonica. La scrittura, registrata su una lavagna o su un pezzo di carta, dovrebbe essere trasformata in un segnale elettrico inviato poi su una linea telefonica. All'altra estremità il laser a elio e cadmio dovrebbe servire come mezzo per trascrivere il messaggio su una piccola diapositiva, da cui lo scritto dovrebbe essere simultaneamente proiettato su uno schermo. Il laser è adatto per questa funzione poiché esso può essere focalizzato in una macchia estremamente piccola sulla pellicola della diapositiva sensibile all'ultravioletto.

Si possono anche immaginare le applicazioni del laser multicolorato a elio e selenio in dispositivi progettati per esaminare e esporre il colore ad alta risoluzione. Inoltre, recenti sviluppi nel campo dei laser a molecole organiche colorate hanno indicato che la potenza emessa da un piccolo laser a elio e selenio può essere sufficiente per rifornire un laser a molecole colorate che operi con continuità. Una tale applicazione dovrebbe condurre a un segnale d'uscita laser poco costoso e armonioso nella regione rossa, arancione e gialla dello spettro visibile.

I pesci a sangue caldo

Non tutti i pesci sono a sangue freddo. Alcuni veloci nuotatori come tonni e ossirine presentano una «rete mirabile» di vene e arterie che mantiene elevata la temperatura dei muscoli del nuoto

di Francis G. Carey

I pesci sono considerati animali a sangue freddo, o piú precisamente animali la cui temperatura corporea è identica a quella dell'ambiente che li circonda. In realtà non tutti i pesci sono a sangue freddo. Il primo a descrivere un pesce a sangue caldo fu un medico inglese, John Davy. Egli, durante un viaggio nei tropici nel 1835, notò la carne molto rossa e riccamente vascolarizzata, come quella dei mammiferi, di uno dei pesci con cui l'alimentazione di bordo veniva variata. Si trattava di una specie di tonno, l'*Euthynnus pelamis.* Egli misurò la temperatura di alcuni di questi pesci e trovò che superava di 10 °C quella delle acque in cui venivano pescati. Successivamente molte altre specie di tonni e anche una famiglia di squali dimostrarono di avere sangue caldo. Recentemente, presso la Woods Hole Oceanographic Institution, insieme ad alcuni colleghi, mi sono dedicato allo studio di queste specie di pesci.

Qual è il meccanismo con cui la maggior parte dei pesci mantiene la propria temperatura del corpo alla stessa temperatura ambientale, nonostante il calore prodotto dal metabolismo? Non si tratta, come si potrebbe supporre, della perdita di calore dovuta al fatto che i pesci si trovano immersi in un mezzo a bassa temperatura. La temperatura del corpo viene invece mantenuta bassa grazie al tipo di respirazione dei pesci. Tutti gli animali terrestri e alcuni animali acquatici ricavano l'ossigeno necessario al loro metabolismo dall'aria, che è ricca di questo gas e assorbe poco il calore. L'acqua, invece, da cui i pesci devono estrarre l'ossigeno contiene solo il 2,5 per cento dell'ossigeno contenuto nello stesso volume d'aria, ma ha una capacità di assorbire calore 3000 volte superiore a quella dell'aria. Da questo fatto si può capire come la perdita di calore che si ha estraendo ossigeno dall'acqua anziché dall'aria sia 100 000 volte superiore, a parità di volume di ossigeno ottenuto.

Nei pesci l'ossigeno contenuto nell'acqua passa attraverso le branchie e si diffonde nel sistema circolatorio; il sangue spinto dal cuore può giungere alle branchie a una temperatura piuttosto elevata a causa del calore metabolico accumulato. Il calore, d'altra parte, si diffonde 10 volte piú rapidamente delle molecole di ossigeno di modo che, quando il sangue nelle branchie è saturo di questo gas, la sua temperatura si è già abbassata fino a raggiungere quella dell'acqua.

Il sangue arterioso ricco d'ossigeno, va a irrorare i tessuti dove questo gas viene utilizzato per il metabolismo, ed è durante questo processo che si ha un aumento della temperatura proporzionale alla quantità di ossigeno che viene estratto dal sangue. In un dato volume di sangue arterioso il volume di ossigeno estraibile è inferiore al 20 per cento. Tenendo conto del normale apporto calorico ricavato dal cibo nei pesci, l'ossigeno necessario a mantenere il metabolismo è poco, e l'aumento della temperatura del sangue non supera 1 °C. Inoltre, anche questo lieve incremento di temperatura viene ceduto all'acqua quando il sangue passa di nuovo attraverso le branchie in modo che la temperatura del pesce si mantiene uguale a quella dell'acqua che lo circonda.

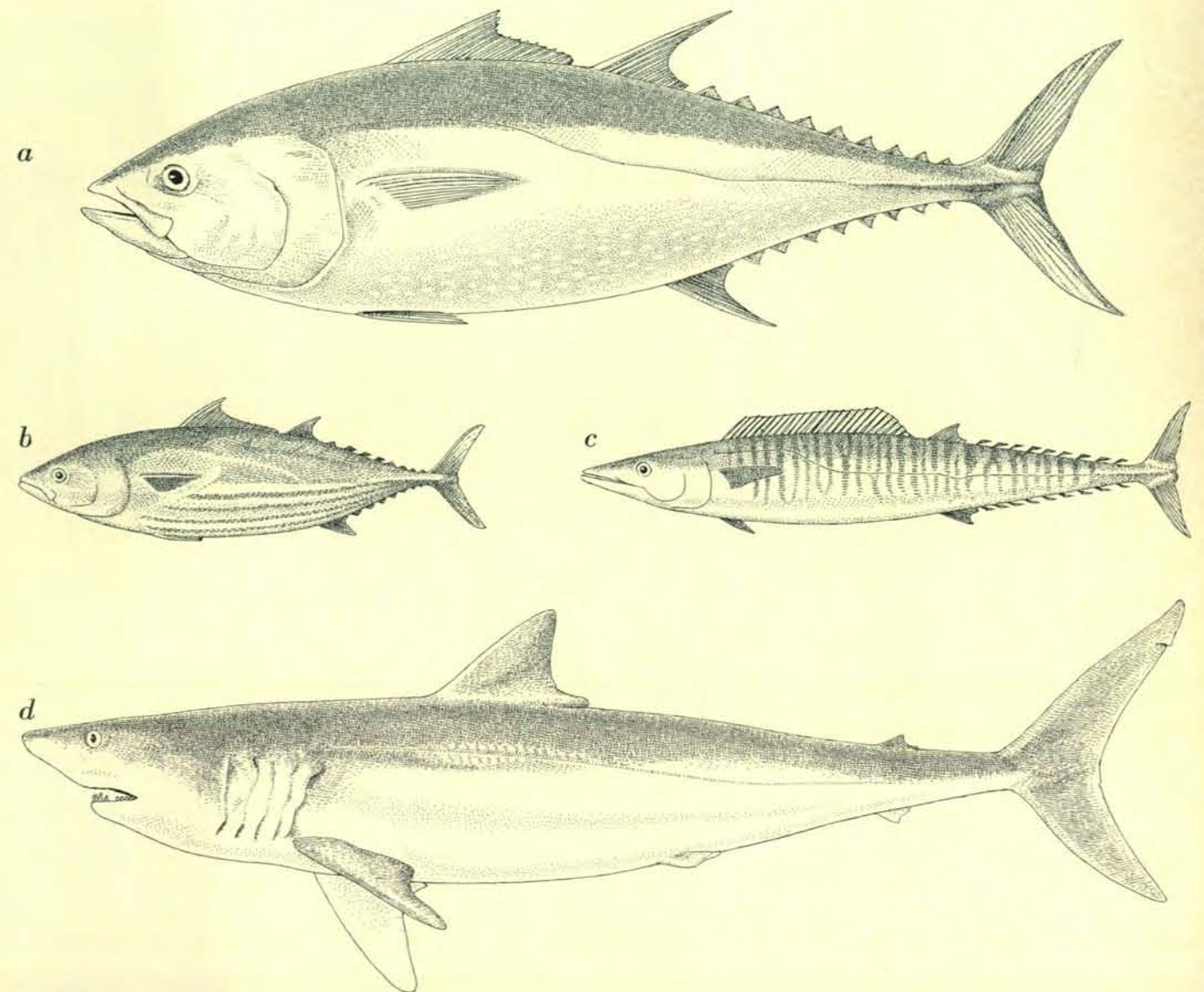
Le variazioni nella velocità del metabolismo non influenzano questo equilibrio in quanto qualunque esercizio o altra attività che produca un sovrappiú di calore metabolico richiede anche piú ossigeno. Una domanda di ossigeno superiore al normale significa una maggiore perdita di calore attraverso le branchie, in modo che il pesce, sia che stia fermo e si riposi, sia che faccia rapidi movimenti, sarà sempre freddo. Sembra dunque impossibile che un pesce riesca ad aumentare la propria temperatura corporea attraverso il metabolismo.

Apparentemente contro questa legge naturale, i tonni e una particolare famiglia di squali, gli isuridi, riescono a mantenersi « caldi », grazie a una particolare struttura del loro sistema circolatorio. Si tratta di una « rete mirabile », cioè di un tessuto costituito da un fitto intreccio di vene e arterie, osservato per la prima volta nei tonni nel 1831 dal naturalista francese Georges Cuvier. Quattro anni piú tardi, il tedesco Johannes Müller fece una descrizione accurata della rete mirabile osservata nei visceri del tonno.

Questa particolare struttura costituisce una barriera termica alla perdita del calore metabolico. L'insieme di venuzze e arteriole permette il passaggio di sangue per il trasporto di ossigeno e di altre molecole, ma contemporaneamente rappresenta un « corto circuito » nei confronti degli scambi di calore tra i tessuti e le branchie e un « by-pass » nei confronti del calore accumulato nei tessuti. Nella rete mirabile un'arteria che trasporta sangue freddo e ben ossigenato proveniente dalle branchie verso i tessuti, si ramifica formando numerosi minuscoli vasi che scorrono parallelamente a un'analoga massa di vasi venosi che trasportano sangue caldo, povero di ossigeno dai tessuti alle branchie. Questo insieme di venuzze e arterie strettamente avvicinate tra loro agisce come uno scambiatore di calore in controcorrente. Tra i due tipi di vasi vi sono infatti vaste superfici di contatto e il calore passa facilmente attraverso le sottili pareti. Il sangue venoso piú caldo viene raffreddato mentre attraversa la rete mirabile, in modo che quando raggiunge le branchie non si ha quasi piú perdita di calore; contemporaneamente il sangue freddo arterioso viene riscaldato in modo che, quando raggiunge i tessuti piú interni del pesce, ne uguaglia quasi la temperatura. Una rete di questo tipo ha esclusivamente la funzione di scambiatore di calore; i suoi vasi, benché di piccolo calibro, sono troppo grandi e le loro pareti sono troppo spesse per permettere gli scambi gassosi tra vene e arterie. In molte specie di pesci esistono altri tipi di reti che hanno la funzione di permettere gli scambi gassosi, ma esse sono formate da vasi capillari attraverso cui le molecole di ossigeno passano facilmente. Queste reti si possono trovare nella vescica natatoria e nell'occhio.

Il vantaggio adattativo di una maggiore temperatura corporea consiste in una maggiore potenza muscolare. Due muscoli identici che abbiano una temperatura che differisce di 10 °C hanno un rendimento assai diverso. Il muscolo piú caldo infatti si contrae e si rilassa tre volte piú rapidamente; questo incremento di potenza è particolarmente utile negli animali che volano come uccelli e pipistrelli, che presentano una temperatura corporea particolarmente elevata. Anche alcuni grandi insetti potenti volatori sono in grado di innalzare notevolmente la loro temperatura durante il volo. Per esempio, cicale e locuste sono spesso incapaci di alzarsi in volo se non si sono prima riscaldate mediante un meccanismo simile al brivido o piú semplicemente esponendosi al sole.

Anche il nuoto veloce, come il volo, è un tipo di locomozione che richiede un notevole dispendio di energia. L'acqua è un mezzo denso e viscoso e, per muoversi rapidamente attraverso di essa, non basta una linea aerodinamica, ma occorrono anche potenti muscoli. L'elevata temperatura del corpo dei tonni e degli isuridi sembra quindi fornire l'energia in piú necessaria al nuoto veloce; si è notato infatti che i piú importanti tra gli scambiatori di calore sopra descritti sembrano essere quelli che servono i muscoli scuri adibiti al nuoto.

I muscoli dei vertebrati sono generalmente costituiti da un insieme di fibre piú scure alternate a fibre piú chiare. Le fibre scure contengono un'elevata concentrazione di pigmenti che trasportano l'ossigeno e di enzimi associati al metabolismo ossidativo; sono inoltre sempre riccamente vascolarizzate. Le fibre piú chiare sono invece meno irrorate di vasi. Esse possono contrarsi an-

I pesci che sono veloci nuotatori e hanno sangue caldo, hanno una forma aerodinamica e una coda molto muscolosa. I pesci dell'esempio sono tre tonni: il tonno comune (*a*), il tonnetto striato (*b*), il wahoo (*c*) e uno squalo: l'ossirina (*d*).

che in assenza di ossigeno metabolizzando gli alimenti attraverso la fermentazione, e pagando quindi il debito energetico contratto nella successiva fase di riposo. I muscoli che sono in continuo movimento, per esempio il cuore, sono formati esclusivamente o in gran parte da fibre scure. I muscoli invece che alternano brevi periodi di lavoro con periodi di riposo non contengono che pochissime fibre scure.

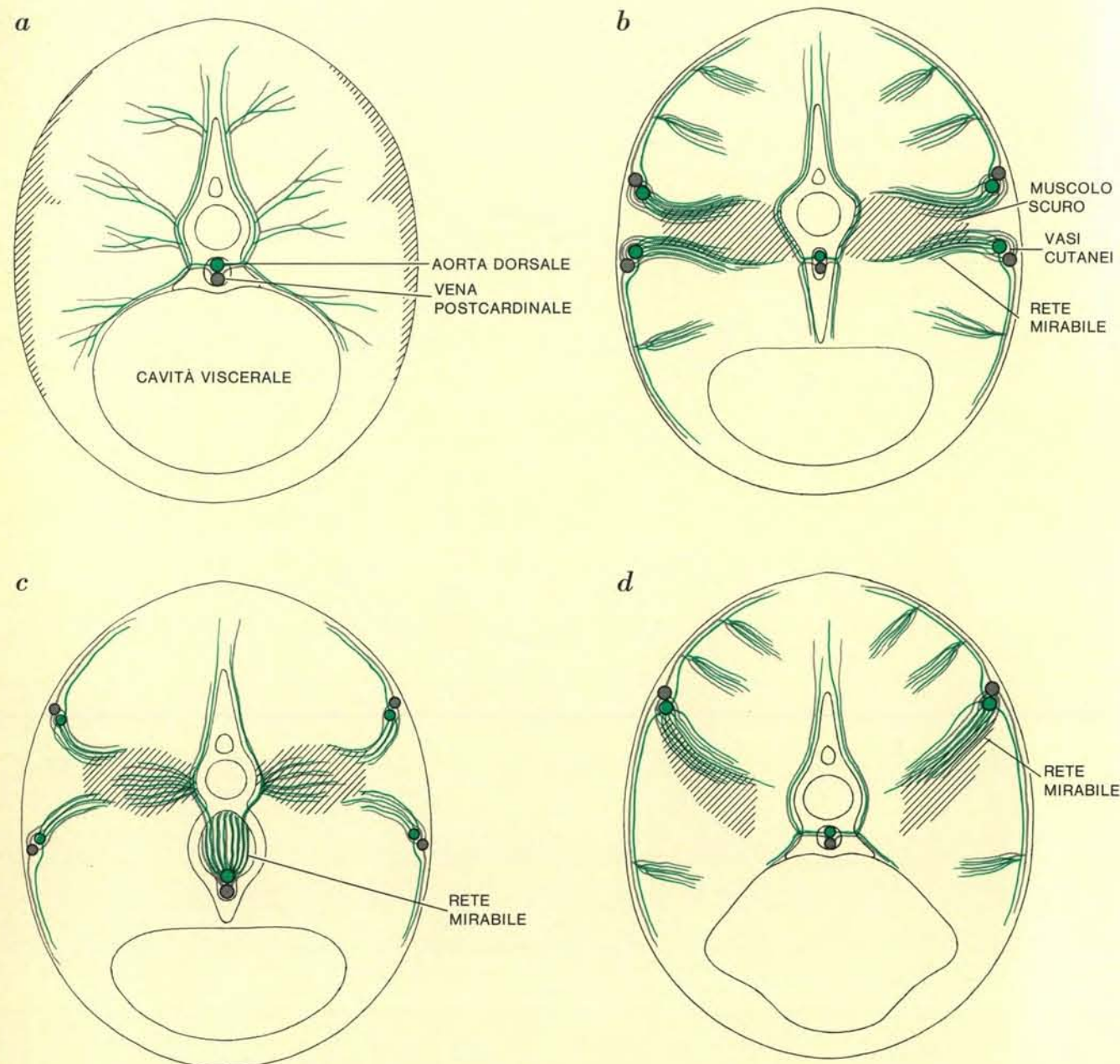
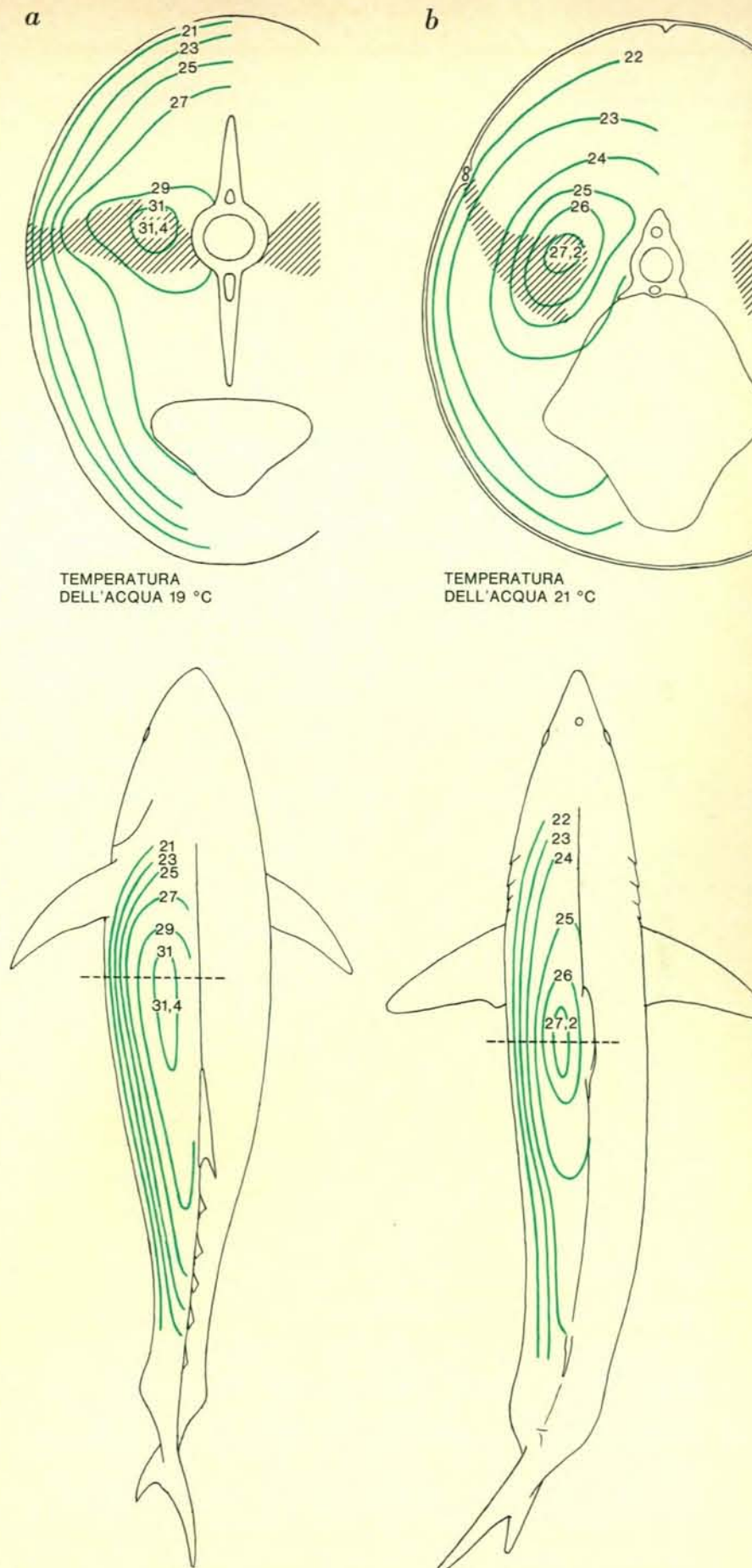
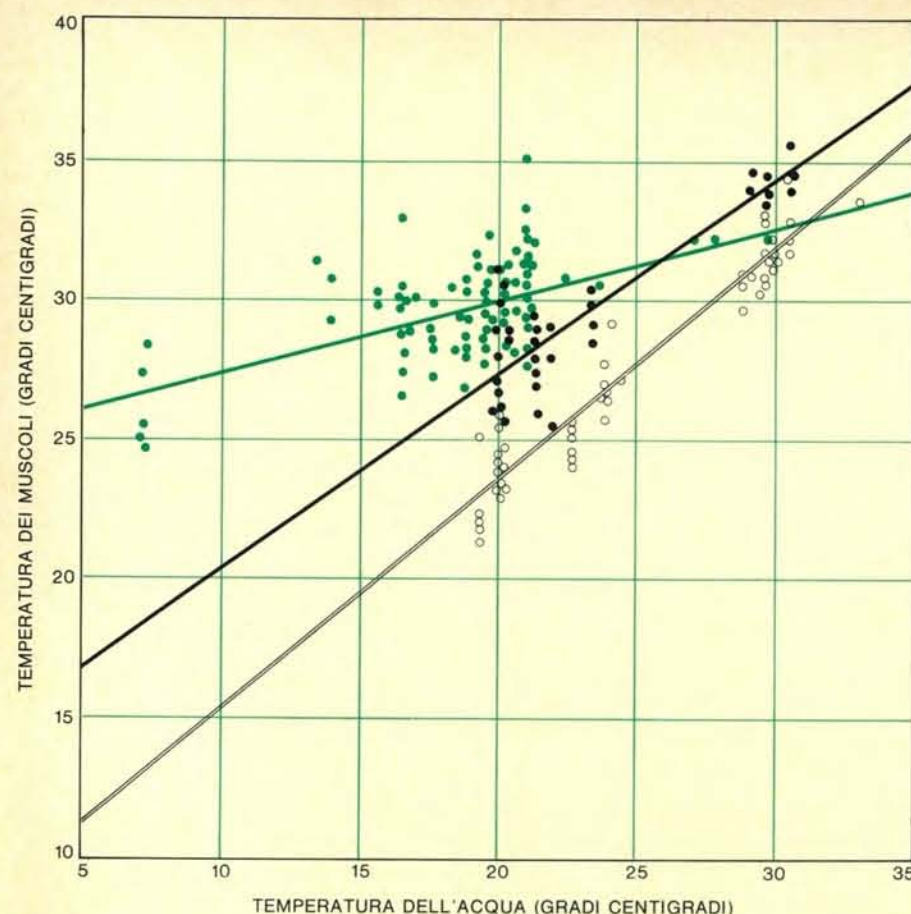
Chiunque abbia visto un trancio di pesce ha osservato che i due tipi di muscolatura sono chiaramente distinguibili nella muscolatura assiale. Per esempio, in molti tonni si può osservare una grande massa di muscolatura molto scura e vascolarizzata situata dietro la colonna vertebrale lungo il dorso del pesce. Questi muscoli possono lavorare in continuazione e le loro contrazioni fanno viaggiare il pesce alla normale velocità di crociera. Il sangue che irrora i muscoli, prima di raggiungere le fibre muscolari, attraversa una ampia rete mirabile; in questo modo questi muscoli hanno una temperatura superiore al resto dell'organismo.

Per poter creare questa ampia rete con funzione di scambiatore di calore, tutto il sistema circolatorio dei tonni ha subito drastiche modifiche. La maggior parte dei pesci presenta un sistema circolatorio « centralizzato »: quasi tutto il sangue che va a irrorare i muscoli passa attraverso una grossa arteria, l'aorta dorsale, e ritorna indietro attraverso la vena postcardinale; entrambi i vasi sono situati in profondità in prossimità della spina dorsale. Da questi vasi centrali si dipartono numerose vene e arterie segmentarie che vanno verso i tessuti periferici.

Nei tonni questo schema circolatorio è completamente capovolto. I vasi principali non sono piú l'arteria e la vena centrali; al loro posto ci sono quattro paia di arterie e di vene sottocutanee, due per ciascun lato del pesce che trasportano la maggior parte del sangue circolante (*si veda l'illustrazione in alto a pagina 52*). Da questi vasi cutanei si dipartono numerosissimi vasi aventi un diametro di appena un decimo di millimetro. Questi minuscoli vasi si mescolano tra loro formando delle reti mirabili situate sopra e sotto le fasce muscolari scure (*si veda l'illustrazione in basso a pagina 52*). Queste reti rappresentano degli scambiatori di calore che assicurano alle fasce muscolari scure una temperatura piuttosto elevata. Anche il sangue che irrora i muscoli piú chiari proviene dai grossi vasi cutanei, attraverso arterie e vene segmentarie che scorrono sopra e sotto la superficie dei muscoli, inviando numerosi rami entro di essi. Questi rami formano delle bande di arterie alternate alle vene, che fungono anch'esse da scambiatori di calore. La superficie di contatto, però, tra arterie e vene in queste strutture bidimensionali è minore di quella della rete mirabile delle fasce muscolari piú scure; è sufficiente tuttavia ad assicurare un adeguato calore a questi tessuti che non hanno un flusso di sangue molto veloce.

Questa particolare distribuzione dei vasi sanguigni principali proprio sotto la pelle presenta il vantaggio di ridurre al minimo la perdita di calore lungo la superficie del corpo. Infatti l'estremità fredda degli scambiatori di calore della rete mirabile è rivolta verso l'esterno, mentre l'estremità calda è situata in profondità nei tessuti.

Abbiamo misurato la temperatura del tonno in varie regioni del corpo riscontrando che le zone piú calde sono quelle delle fasce muscolari scure (*si veda l'illustrazione a destra*). La temperatura piú elevata però non si trova all'interno dei muscoli. Questo perché le regioni piú interne vengono in parte irrorate dall'aorta dorsale attraverso dei vasi che non si sono specializzati come scambiatori di calore. La temperatura piú alta nei muscoli è stata misurata in prossimità della zona centrale della fascia muscolare che scorre lungo i fianchi del pesce, ma in generale quasi tutti i muscoli, sia chiari che scuri, presentano una temperatura notevolmente superiore a quella dell'acqua in cui il pesce si trova.

Il tonno alalunga ha una circolazione simile a quella del tonno comune. Altre specie di tonni, in particolare il tonnetto striato e un po' meno il

In un pesce a sangue freddo (*a*) i vasi sanguigni principali si originano dai grossi vasi centrali. Nei pesci a sangue caldo il sistema circolatorio ha una disposizione completamente diversa per far sí che i muscoli scuri ricevano sangue preventivamente riscaldato. Nel tonno (*b*) esiste una rete mirabile che forma delle bande di tessuto vascolare sopra e sotto il muscolo. Nel tonnetto striato (*c*), invece, la rete è formata da un ammasso di piccoli vasi che scorrono verticalmente sotto alla spina dorsale. Nell'ossirina (*d*) esiste un solo paio di vasi cutanei, che alimentano una massiccia rete situata su ciascun lato del pesce.

Temperature riscontrate in varie regioni del corpo nel tonno (*a*) e nell'ossirina (*b*). Le isoterme indicano che le zone piú calde si trovano entro i muscoli scuri lungo i fianchi del pesce. Anche gli altri muscoli hanno una temperatura piú elevata dell'acqua.

Il grafico mette in relazione la temperatura dell'acqua con quella massima riscontrata nei muscoli del tonno (*in colore*), del tonnetto striato (*in nero*) e del tonno pinna gialla (*doppia linea*). Il tonno regola la sua temperatura abbastanza indipendentemente dall'acqua in cui si trova. Il tonnetto striato è piú caldo del tonno pinna gialla, ma entrambi tendono a mantenere una temperatura costante superiore a quella dell'acqua.

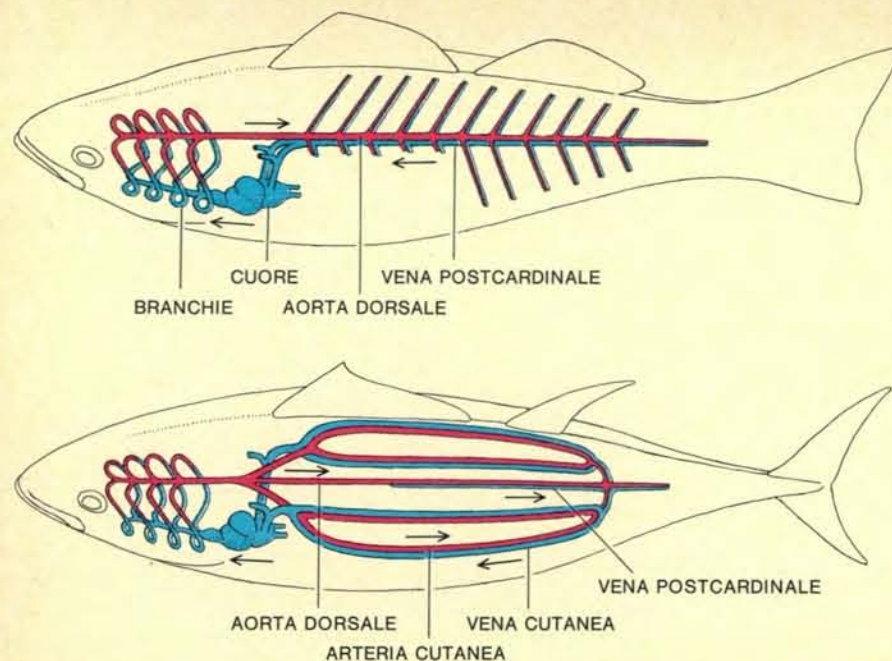
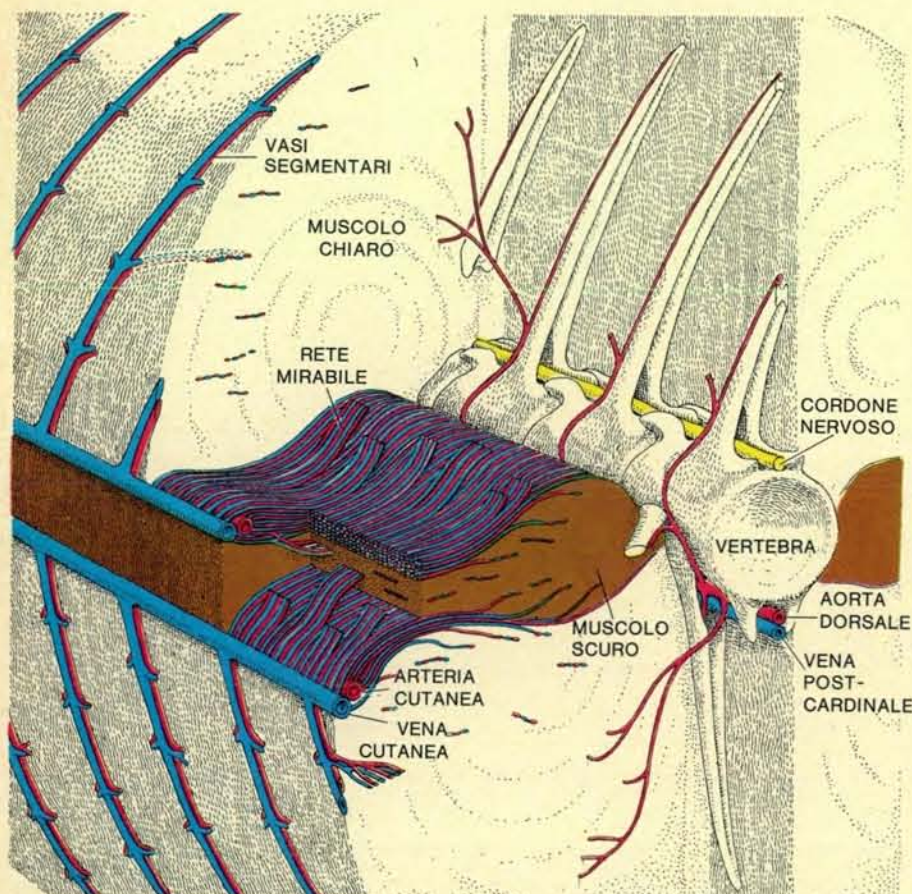
tonno pinnagialla possono innalzare la temperatura del loro corpo con sistemi diversi. Benché anche in essi siano presenti le arterie e le vene cutanee, gli scambiatori di calore a cui danno luogo sono formati da un singolo strato di sottili vasi sanguigni. Lo scambiatore di calore principale è invece collegato con l'arteria e la vena centrali situate sotto la colonna vertebrale. Questa rete centrale forma una massa cilindrica piú larga della spina dorsale e che si estende lungo le vertebre (*si veda l'illustrazione a pagina 48*). Le misurazioni della temperatura corporea hanno dimostrato la grande efficienza di questo tipo di struttura: questi piccoli pesci hanno infatti una temperatura che spesso supera quella dell'acqua di 10 °C e i loro muscoli piú caldi sono situati lungo la spina dorsale.

È interessante notare come anche una famiglia di squali, gli isuridi che si sono evoluti assai piú anticamente dei pesci ossei e in particolare di quelli piú specializzati, come i tonni, abbiano sviluppato un sistema scambiatore di calore molto simile. I vasi sanguigni che scorrono in prossimità della colonna vertebrale sono piccoli, mentre la maggior parte del flusso sanguigno passa attraverso una sola coppia di vasi (vena e arteria) che scorre proprio sotto la pelle lungo i fianchi del pesce. Questo paio di vasi cutanei alimenta una sola massiccia rete che mantiene calde le fasce muscolari scure dello squalo. In alcuni di questi squali, per esempio nell'ossirina, la rete mirabile forma uno strato compatto, proprio come nel tonno (*si veda l'illustrazione a pagina 48*). In altri, come nello squalo bianco, la rete è piú estesa con numerosi fasci di sottili vasi sanguigni che si insinuano tra i muscoli. Negli isuridi la distribuzione della temperatura del corpo è molto simile a quella del tonno; la maggior parte dei muscoli dello squalo ha una temperatura piú elevata di quella dell'acqua che lo circonda e le zone piú calde sono due muscoli gemelli posti entro la muscolatura scura.

I tonni e gli isuridi possiedono inoltre scambiatori di calore che non sono associati alla muscolatura, ma servono gli organi posti nella cavità del corpo. Nei tonni si possono osservare densi ammassi di sottili vasi sanguigni che rivestono la superficie del fegato; in alcuni esemplari particolarmente grandi questi fasci possono avere un diametro di cinque centimetri. Negli isuridi si ha invece una disposizione differente, in quanto l'arteria, che negli altri pesci porta il sangue ai visceri, è piccolissima o addirittura inesistente. Invece vi sono altre arterie, di scarsa importanza negli altri pesci, che irrorano di sangue i visceri. Queste arterie si immettono in un ampio spazio chiamato seno epatico. Il sangue venoso che va verso il cuore attraversa questo seno. Le arterie formano numerosissimi rami entro il seno epatico dando luogo a una specie di massa spugnosa formata da sottilissimi vasi; il sangue freddo in essi contenuto può quindi scaldarsi a contatto con il piú caldo sangue venoso.

Questa rete viscerale non funziona in continuazione, la temperatura dei visceri può variare tra quella dei muscoli piú caldi e una temperatura solo di poco superiore a quella dell'acqua. Queste variazioni sono correlate con la attività digestiva, in particolare per i tonni che hanno una cavità viscerale ristretta e apparato digerente piccolo. In questo caso un innalzamento della temperatura dei visceri può essere utile per accelerare la digestione. Quando questo non è necessario, i visceri possono anche non venire riscaldati.

Dal momento che i tonni sono a sangue caldo, ci si può chiedere quale sia il meccanismo con cui regolano la loro temperatura e la mantengono a un livello costante indipendentemente da quella dell'acqua che li circonda che può variare di molto. Altri ricercatori hanno notato che il tonno alalunga e il tonnetto striato, tendono a mantenere la loro temperatura superiore a quella dell'acqua intorno a un dato nu-

Sezione trasversale della rete mirabile di un tonno. La rete è un sistema costituito da piccole arterie e vene che irrorano le fasce muscolari scure dei pesci che vengono utilizzate nel nuoto. Il sistema ha la funzione di un vero e proprio scambiatore di calore in controcorrente: il sangue venoso che viene scaldato dal metabolismo cellulare cede il suo calore al sangue arterioso piú freddo perché proveniente dalle branchie. Nella fotografia della pagina a fronte abbiamo distinto nella rete mirabile del tonno i vasi venosi da quelli arteriosi iniettando lattice blu nei primi e rosso nei secondi.

Confronto tra il sistema circolatorio di un pesce a sangue freddo (*in alto*) e a sangue caldo (*in basso*). Nei pesci a sangue freddo i vasi principali scorrono lungo la spina dorsale e si irradiano verso i vasi minori (*non rappresentati*) che irrorano i muscoli. Nei pesci a sangue caldo i vasi centrali hanno scarsa importanza; la maggior parte del sangue scorre infatti attraverso i vasi sottocutanei e gli ammassi di piccoli vasi che funzionano da scambiatori di calore e che portano il sangue ai muscoli. Questo tipo di organizzazione fa sí che le regioni apicali fredde degli scambiatori siano rivolte verso la cute, mentre quelle terminali calde sono rivolte verso i tessuti interni.

Sono essenzialmente quattro paia i vasi cutanei che trasportano la maggior parte di sangue nel tonno. Da essi si ramificano innumerevoli minuscoli vasi che formano un tessuto che irrora e riscalda la faccia superiore e inferiore delle fasce muscolari scure. I muscoli chiari ricevono l'apporto di arterie e vene alternate che provengono da vasi segmentari e hanno anch'essi la funzione di scambiatori di calore a due dimensioni. Dai vasi centrali si originano anche numerose altre vene e arterie.

mero di gradi (questa regolazione potrebbe essere mantenuta da un sistema anche non perfetto di scambiatori di calore). Bisogna però tenere conto del fatto che queste due specie vivono in acque la cui temperatura è quasi costante; ci è sembrato interessante quindi prendere in considerazione il tonno comune che vive sia nelle acque tropicali che in quelle prossime al polo. Misurammo la massima temperatura dei muscoli di questo tonno in esemplari presi lungo le coste che vanno dalle Bahamas a Newfoundland. Il confronto tra la temperatura dei muscoli e quella delle acque in cui i tonni erano stati pescati, dimostra che questi pesci sono in grado di controllare abbastanza bene la loro temperatura corporea. Quando la temperatura dell'acqua è intorno ai 24 °C, la temperatura dei muscoli aumenta di soli 6 °C; i tonni delle Bahamas, le cui acque hanno una temperatura di 30 °C, erano solo di pochi gradi piú caldi dell'acqua, mentre quelli pescati in acque a 7° di latitudine piú a nord avevano una temperatura che superava di 20 °C quella dell'acqua.

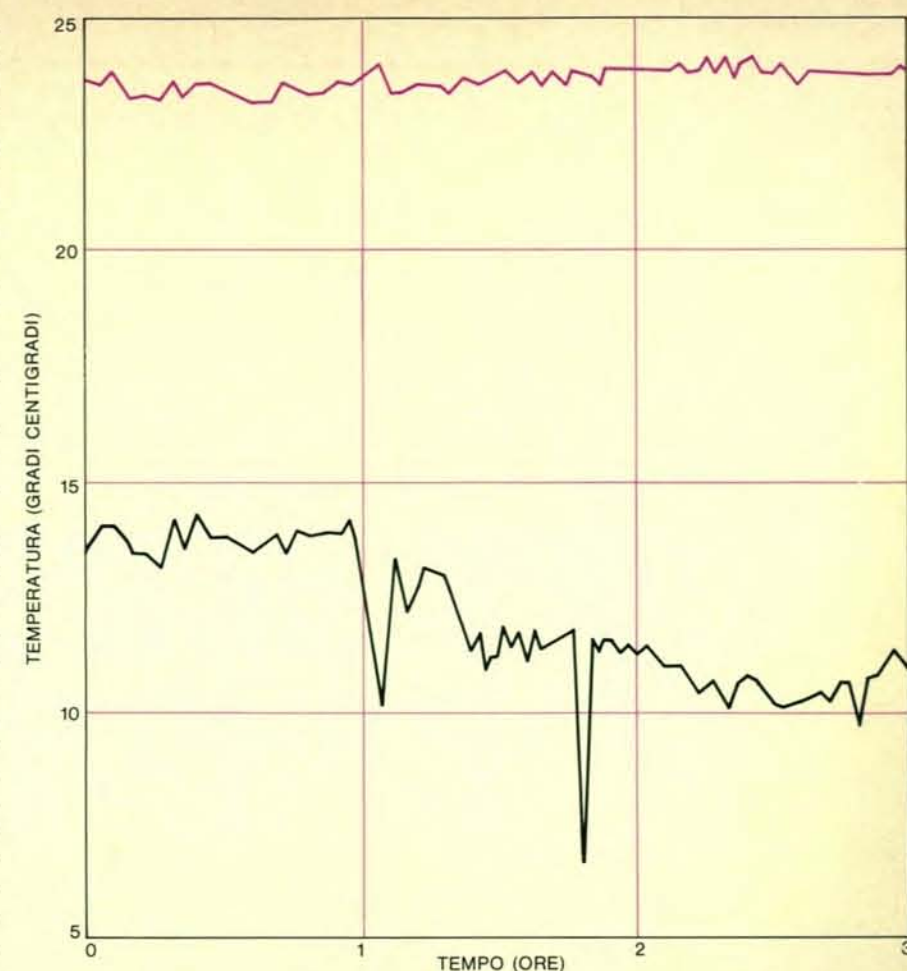
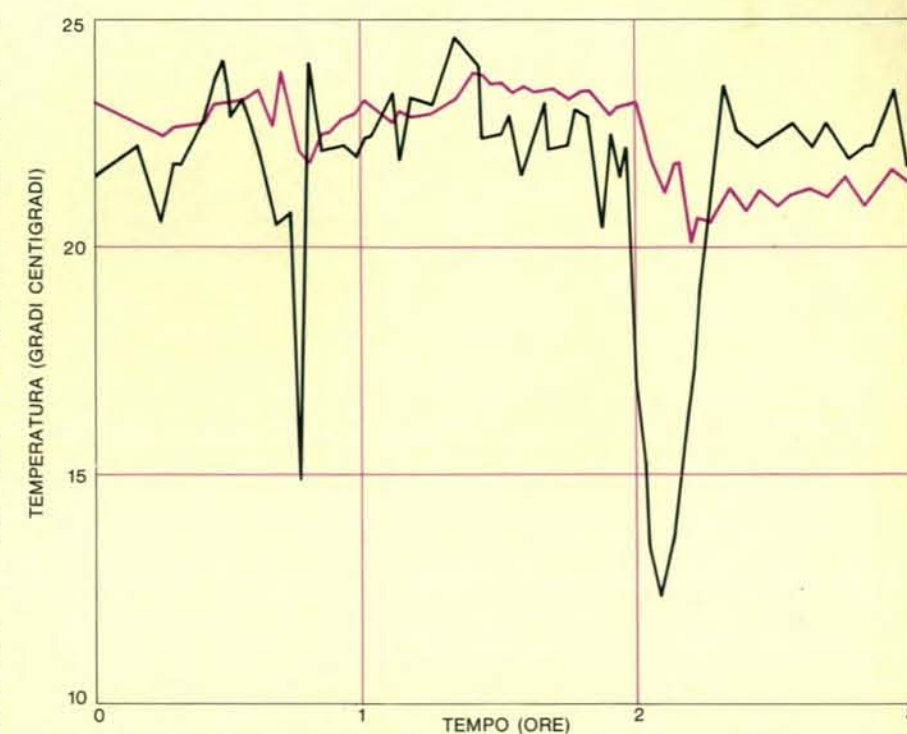
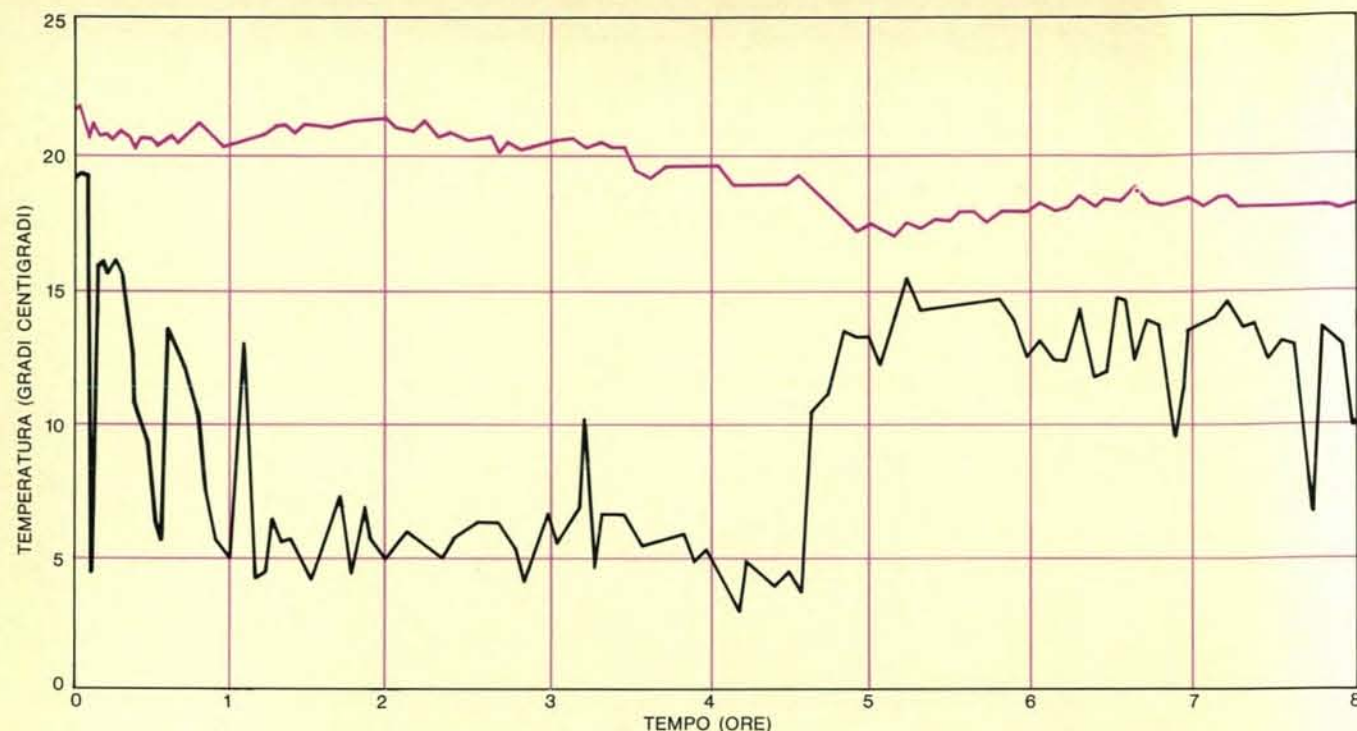
Gli animali hanno due modi per controllare la propria temperatura: possono avere un apparato termoregolatore che si adatta rapidamente alle variazioni ambientali come quello di uccelli e mammiferi che permette loro di variare la velocità con cui perdono calore in vari modi e di produrre calore in sovrappiú aumentando il metabolismo. Gli animali a sangue freddo hanno invece un processo di acclimatazione piú lento che richiede che essi adattino la loro attività e che coinvolge cambiamenti cellulari ed enzimatici, che avvengono in un periodo di tempo piuttosto lungo. Per scoprire quale fosse il meccanismo usato dai tonni, decidemmo di tentare un esperimento da eseguire in mare, esperimento in cui avremmo potuto registrare la temperatura del pesce mentre si spostava da acque fredde ad acque calde.

La costa della Nuova Scozia si dimostrò il luogo ideale per la realizzazione del nostro progetto: vi lavora un gruppo di pescatori che stendono le loro reti nella baia di St. Margaret prendendo spesso anche dei tonni; inoltre il mare presenta una netta termoclina (un rapido abbassamento della temperatura con l'aumento della profondità) per cui un pesce libero di nuotare può attraversare acque a temperature molto diverse tra loro.

Usammo un piccolo arpione fornito di due termistori per misurare la temperatura dei tonni. Un termistore misurava la temperatura dell'acqua e un altro quella della muscolatura del pesce; essi erano collegati a un indicatore mediante un cavo telefonico lungo circa 1 chilometro che poteva svolgersi quando il pesce si allontanava nuotando. In questi nostri primi tentativi non ottenemmo praticamente alcun risultato: o il pesce moriva o l'arpione si staccava quando il cavo si impigliava nelle alghe o nelle reti. A questo punto il mio collega John Kanwisher pensò di effettuare i nostri rilevamenti per mezzo di un sistema di telemisura.

A questo scopo scegliemmo dei piccoli trasmettitori acustici alimentati a batteria, in grado di trasmettere i loro segnali fino a circa otto chilometri, e dotati di un'autonomia di qualche giorno. Il trasmettitore veniva montato su un piccolo arpione che poteva venire inserito nel muscolo del pesce provocando una ferita trascurabile. Sulla punta dell'arpione era montato il termistore per il rilevamento della temperatura del muscolo, mentre quello per la misura della temperatura dell'acqua era sistemato sul trasmettitore, fuori del corpo del pesce. Per altri esperimenti, nei quali misuravamo la temperatura dello stomaco, il trasmettitore con il relativo termistore venivano fatti ingoiare dall'animale, mentre il termistore per la misura della temperatura dell'acqua, questa volta, era montato al termine di un cavo che veniva fatto uscire attraverso le aperture branchiali. Questi trasmettitori davano la misura della temperatura interna e di quella dell'acqua alternativamente, per periodi di circa 1 minuto l'una.

L'imbarcazione con la quale seguivamo i nostri pesci era munita di un idrofono direzionale, mediante il quale si individuava la direzione di massima intensità del segnale della trasmittente, localizzandone cosí la posizione, in modo da poter rimanere nelle sue vicinanze. A intervalli di un'ora immergevamo un batitermometro per rilevare la temperatura dell'acqua a varie profondità. Questo ci consentiva di risalire alla quota alla quale nuotava il pesce in quel momento, dato che la trasmittente ci inviava la temperatura dell'acqua in prossimità dell'animale. Il percorso compiuto dall'animale era poi quello seguito dall'imbarcazione, se questa riusciva a mantenersi sulla sua verticale mentre si ricevevano i segnali della trasmittente.

Riuscimmo a installare le nostre trasmittenti su 14 tonni. Una volta riuscimmo a tenere il contatto col pesce che nel frattempo si era spostato a 130 miglia dalla costa, per 54 ore. Presto fu possibile stabilire che, come ci era stato detto dai pescatori, i tonni evitano, se possibile, di portarsi in acque a temperatura molto diversa da quella

Le registrazioni telemetriche permettono di confrontare la temperatura del muscolo del tonno (*in colore*) con quella dell'acqua (*in nero*). Durante la prova, la temperatura del muscolo si mantenne costante mentre quella dell'acqua si abbassò da 14 a 10 °C.

Registrazioni della temperatura del muscolo di un pescecane (*in colore*) e dell'acqua in cui nuota (*in nero*). Il pescecane ha sangue freddo e la temperatura dei suoi muscoli, benché si mantenga quasi costante, è molto vicina a quella dell'acqua.

Il differente sistema di controllo della temperatura dei visceri rispetto a quella dei muscoli è messo in evidenza da queste registrazioni della temperatura dello stomaco di un tonno. La temperatura venne registrata telemetricamente mediante un termistore posto nello stomaco del tonno (*in colore*); contemporaneamente veniva trasmessa anche la temperatura dell'acqua (*in nero*). Appena liberato il tonno andò a rifugiarsi in acque fredde, ma la sua temperatura interna diminuí solo di poco.

in cui si trovano abitualmente. La maggior parte dei nostri esemplari rimaneva vicino alla superficie, o almeno sul lato caldo del ripido gradiente di temperatura che separava le acque superficiali calde da quelle fredde piú profonde. Alcuni tonni si immergevano attraverso il gradiente, ma rimanevano nelle acque fredde solo per pochi minuti, tornando al piú presto negli strati caldi.

Alla fine del nostro lavoro avemmo la fortuna di ottenere un risultato estremamente interessante con un tonno di circa tre quintali a cui avevamo posto una trasmittente nello stomaco. Il pesce era stato piuttosto maltrattato per fargli ingoiare l'apparecchio e, forse per questo motivo, si inabissò rapidamente appena liberato, nuotando in acque di 14 gradi piú fredde di quelle superficiali in cui era stato catturato. All'inizio dell'esperimento la temperatura dello stomaco era di 21 °C. Nel corso della permanenza in acque fredde, durata quattro ore, la temperatura dello stomaco scese gradualmente a 19 °C. Quando il pesce tornò in acque calde, rimanendo per il resto della giornata a temperature tra i 13 e i 14 °C, la temperatura dello stomaco rimase stazionaria intorno ai 18 °C, malgrado la temperatura ambiente fosse salita di circa 10 °C. Il fatto che la temperatura delle zone profonde del corpo sia rimasta pressoché costante per lunghi periodi in ambienti caldi e freddi è una prova che quel tonno stava davvero termoregolandosi. Con quale meccanismo, per ora non lo sappiamo. È possibile che riesca in qualche modo a variare l'efficienza di scambio termico delle sue reti mirabili.

Nel corso delle misure di temperatura muscolare non abbiamo avuto la fortuna di imbatterci in un tonno che rimanesse a lungo in acque fredde, come nel caso precedente. Però uno dei nostri tonni nuotò per varie ore in acque che diminuivano lentamente di temperatura. Questa scese di quattro gradi nel corso di un'ora e, nel corso di questo tempo, la temperatura muscolare del pesce aumentò lentamente da poco piú di 23 °C fino a oltre 24.

I nostri esperimenti hanno dimostrato che i tonni non mantengono la loro temperatura corporea a un livello costante, come fanno, per esempio, gli esseri umani. La temperatura dei pesci di uno stesso branco può variare anche di cinque gradi tra un esemplare e l'altro. La maggiore oscillazione registrata in uno stesso pesce è stata osservata in un tonno che portava un trasmettitore di temperatura nello stomaco. Appena rimesso in libertà, questo pesce aveva una temperatura di 19 °C, che cominciò a salire lentamente durante il primo giorno di registrazione. Il secondo giorno, benché la temperatura dell'acqua fosse rimasta praticamente costante, nello stomaco del tonno si registrava una temperatura di 26 gradi, ma forse il pesce stava cercando di digerire la trasmittente!

In altri casi abbiamo osservato che la temperatura muscolare oscillava entro un ambito ristretto, non correlato con la temperatura dell'acqua. Tutta questa variabilità di comportamento sta a indicare che, anche se la temperatura degli organi interni del tonno è molto maggiore di quella dell'acqua, una termoregolazione precisa non è necessaria o possibile. In effetti la capacità di termoregolazione dei pesci è piuttosto scarsa. Ciononostante si può dire che questi pesci a sangue caldo termoregolano, proprio come gli uccelli o i mammiferi.

Come si è sviluppato questo meccanismo di scambio termico nei tonni e negli isuridi? Per quanto la rete mirabile sia un sistema vascolare indubbiamente complesso, la sua evoluzione è stata probabilmente assai semplice. Innanzitutto le vene e le arterie seguono praticamente il medesimo percorso all'interno degli organi e dei tessuti. Questo accoppiamento costituisce di per sé un rudimentale sistema di scambio termico e, se vi è una differenza di temperatura tra le due estremità di esso, si registra pure un passaggio di calore tra i due vasi. Questo rudimentale sistema di scambio è probabilmente responsabile del fatto che, anche i pesci che non sono in grado di aumentare la loro temperatura rispetto all'ambiente utilizzando la produzione di calore dovuta al metabolismo, hanno in realtà una temperatura superiore di uno o due gradi a quella dell'acqua.

Poniamo che questo piccolo aumento di temperatura provochi un qualche vantaggio. Se è cosí, già nel normale sviluppo embrionale del sistema circolatorio esiste il presupposto per la formazione di un sistema di scambio termico piú progredito di quello dei vasi accoppiati. Infatti il sistema circolatorio embrionale non è suddiviso inizialmente in vasi distinti, ma è costituito da varie reti di spazi e canali intercomunicanti. Successivamente alcuni canali piú importanti si accrescono, mentre altri si atrofizzano e scompaiono. Inoltre si osserva una certa tendenza allo sviluppo di canali multipli. Questi possono formarsi in un certo numero all'interno di una delle reti interconnesse embrionali, col risultato che il sistema vascolare dell'esemplare adulto possiede alcuni componenti in doppio. Ecco quindi che, per formare una rete di vasi a decorso parallelo che funzionino da scambiatore di calore, non è necessaria una mutazione genetica radicale, ma solo una leggera variazione nel normale sviluppo embrionale.

Un pesce predatore molto veloce può trovare cibo abbondante sotto forma di calamari, aringhe e sgombri, troppo veloci per i predatori meno scattanti. Ora, in termini di potere di penetrazione nell'acqua e di potenza muscolare, i tonni sono probabilmente i piú efficienti predatori d'alto mare. Il tonno alalunga può raggiungere la velocità di oltre 70 km all'ora in scatti della durata di 10 o 20 secondi. Anche gli squali isuridi, almeno l'ossirina e la *Lamna cornubica*, sono probabilmente avvantaggiati dalla loro abilità nel nuoto veloce. Questi squali, con il loro corpo massiccio e muscoloso, la loro forma penetrante e la pinna caudale sottile e a forma di mezzaluna, assomigliano notevolmente ai tonni. Il fatto che questi due pesci cosí lontani sul piano della filogenesi abbiano sviluppato indipendentemente il medesimo sistema per alzare la temperatura corporea, dipende sicuramente dai vantaggi offerti dalla velocità nel nuoto. Il sovrappiú di potenza ottenibile da un muscolo a temperatura elevata deve essere stato determinante nell'aumentare la velocità del nuoto. È un esempio classico di convergenza evolutiva che ha consentito a entrambe le specie l'accesso a una sorgente di cibo poco sfruttata.

L'origine della specificità delle cellule nervose

I singoli neuroni differiscono l'uno dall'altro più di quanto avviene per le cellule degli altri tessuti. Ogni giovane neurone sembra possedere un «indirizzo» specifico nello stabilire rapporti con altre cellule nervose

di Marcus Jacobson e R. Kevin Hunt

Lo sviluppo del cervello nell'embrione è un avvenimento perfettamente coordinato il cui risultato si trova scritto nei geni, e le cui fasi successive sono prevedibili e immutabili al massimo grado. I neuroni, o cellule nervose, nascono e si sviluppano per svolgere i loro diversi ruoli in un ordine spaziale e temporale regolare. I giovani neuroni prolungano le fibre – gli assoni e i dendriti – che si estendono in direzioni definite per venire a contatto con altre cellule nervose. Alcuni di questi contatti si sviluppano in sinapsi stabili che sono punti di comunicazione tra i neuroni. In questo modo i principali circuiti neuronali raggiungono un'architettura talmente complessa da togliere il respiro, ma semplice nella sua variabilità: le posizioni delle cellule nervose, le traiettorie delle fibre nervose e le disposizioni spaziali delle connessioni sinaptiche sono invariabili in tutti gli individui della stessa specie. Questa invariabilità è detta specificità neuronale.

Lo sviluppo ordinato dei circuiti cerebrali dipende dalla differenziazione di cellule con proprietà specifiche in posizioni particolari e dalla formazione di connessioni ordinate spazialmente tra le cellule. Il significato di una siffatta disposizione spaziale nella differenziazione delle cellule nervose è che la singola cellula in sviluppo acquista alcune proprietà che dipendono dalla sua posizione nei confronti delle altre cellule. Consideriamo un esempio specifico: le cellule del ganglio retinico dell'occhio di rana. Queste differiscono dagli altri neuroni retinici in quanto occupano uno strato distinto della retina e hanno dei lunghi assoni che insieme costituiscono il nervo ottico che connette la retina al tetto ottico e ad altri centri visivi del cervello.

Le singole cellule gangliari si connettono selettivamente con i neuroni nel tetto ottico del cervello in modo tale che la topografia delle posizioni delle cellule gangliari nella retina sia speculare alla topografia delle loro connessioni nel tetto ottico. Una domanda ovvia è perciò in quale modo le cellule gangliari si differenzino dagli altri tipi di neuroni e, sul piano di una differenziazione più particolareggiata, differiscano le une dalle altre. Una seconda domanda è in qual modo gli assoni di queste cellule siano in grado di scegliere il cammino dall'occhio ai centri visivi del cervello. Una terza domanda è in qual modo gli assoni, le singole fibre del nervo ottico, siano collegate mediante connessioni sinaptiche con particolari tipi di cellule nervose in posizioni specifiche nei centri visivi. Sebbene a ogni stadio operino dei meccanismi diversi, il risultato finale è la formazione di disposizioni spazialmente ordinate di cellule nervose, assoni e connessioni sinaptiche. Ciò significa che la cellula nervosa dimostra, in ogni stadio dello sviluppo, di essere strettamente dipendente dalla posizione in cui si trova.

La maggior parte di questi concetti generali trae origine dal lavoro di Roger Sperry dell'Istituto di Tecnologia della California, che compí degli studi, divenuti ormai classici, sulla rigenerazione del nervo ottico di rana. Egli, circa venti anni fa, formulò l'ipotesi che deve esistere uno straordinario grado di differenziazione chimica tra le cellule del sistema nervoso. Sperry scriveva: « Le fibre ottiche differiscono le une dalle altre qualitativamente a seconda del luogo esatto della retina in cui si trovano le cellule gangliari. La retina è apparentemente sottoposta, durante lo sviluppo, a una differenziazione polarizzata che rende localmente specifiche le cellule gangliari e i loro assoni ottici. Le relazioni funzionali stabilite a opera delle fibre ottiche nei centri cerebrali sono disposte in modo sistematico sulla base di questa specificità retinica ».

Nel nostro lavoro all'Università Johns Hopkins e all'Università della Pennsylvania ci siamo particolarmente interessati a quella che definiamo « specificità di luogo »: la proprietà, dipendente dalla posizione che ha ogni cellula gangliare retinica di raggiungere il luogo esatto nella mappa di proiezione delle fibre del nervo ottico nel tetto. Quello che noi vogliamo sapere è in qual modo e quando la dipendenza dalla posizione viene determinata nella retina dell'embrione. Riguarda una sola caratteristica chimica su ogni cellula gangliare, come Sperry ipotizzò per primo? Il nostro lavoro consiste nell'esaminare i cambiamenti di stato attraverso i quali passa la popolazione di cellule retiniche man mano che le cellule gangliari sviluppano la loro organica disposizione di specificità di luogo. Abbiamo scelto per questo lavoro la rana africana *Xenopus laevis*. Il suo sistema nervoso è costruito sullo stesso piano generale di quello degli altri vertebrati, uomo compreso, ma nella rana i centri visivi dell'embrione sono accessibili alle operazioni chirurgiche, gli occhi sono facilmente trapiantabili da un embrione all'altro e i trapianti non vengono rigettati.

Un ulteriore vantaggio dello *Xenopus* è la perfetta conoscenza delle varie fasi del suo sviluppo anatomico. L'occhio origina da una evaginazione del cervello che raggiunge lo strato esterno dell'embrione allo stadio embrionale 22 e quindi gradualmente si ripiega per formare il calice ottico (*si veda l'illustrazione a pagina 59*). Le prime cellule gangliari si sviluppano tra gli stadi 29 e 31, per essere seguiti subito dopo da altri tipi di neuroni retinici. Le cellule disposte intorno alla retina continuano a dividersi e danno luogo ad altri tipi di neuroni retinici durante lo sviluppo, in modo che i neuroni piú vecchi si trovano piú vicino al centro della retina e i piú giovani alla sua periferia.

Ogni cellula gangliare, poche ore dopo la sua formazione, comincia a emettere un assone che raggiunge il tetto ottico dopo circa quindici ore. Inserendo un microelettrodo direttamente nella regione terminale di questi assoni ottici, possiamo stabilire la loro distribuzione spaziale nel tetto del girino o, piú tardi, dell'animale adulto. Noi otteniamo una mappa tetto-retinica rilevando punto per punto la corrispondenza tra la posizione di uno stimolo visivo offerto all'occhio e la posizione della risultante risposta elettrica registrata nel tetto.

Gli esperimenti da noi eseguiti consistevano nell'operare chirurgicamente sull'occhio dell'embrione e poi stabilire la posizione tetto-retinica dopo la metamorfosi del girino in un giovane adulto. Il primo problema era il momento in cui operare; a quale stadio di sviluppo l'organismo era in grado di fornire delle informazioni utili sui processi che volevamo studiare? Circa trenta anni fa, Leon Stone dell'Università di Yale dimostrò l'esistenza di un breve periodo critico nei primi stadi embrionali durante il quale avveniva un cambiamento che interessava il destino dell'intero sistema tetto-retinico. Egli trovò che se l'occhio di un tritone o di una salamandra veniva ruotato prima dello stadio critico embrionale, l'animale sviluppava una visione normale; se l'occhio veniva ruotato in uno stadio piú avanzato, l'animale si comportava come se vedesse il mondo capovolto, e il suo comportamento visivo capovolto non veniva mai corretto dalla pratica. Circa sette anni fa uno di noi (Jacobson) confermò le osservazioni di Stone e dimostrò che il periodo critico corrispondente nello *Xenopus* era tra gli stadi embrionali 28 e 32.

Una rana con tre occhi serve per stabilire in quale misura il terzo occhio che è stato trapiantato nell'embrione con diverse modalità, abbia raggiunto un grado di « specificità » tale da regolare le sue relazioni con le cellule dei centri visivi del cervello. L'animale da esperimento è lo *Xenopus laevis*, nel quale l'occhio può essere facilmente operato e trapiantato nell'embrione.

La mappa tetto-retinica, stabilita con sonde che potevano rivelare l'attività elettrica delle cellule nervose, era normale nello *Xenopus* adulto se l'occhio era stato ruotato di 180° allo stadio 28. Quando la stessa operazione veniva eseguita poche ore dopo, allo stadio 30, la mappa tetto-retinica era capovolta lungo un asse. Le fibre ottiche occupavano le loro posizioni normali nell'asse medio-laterale (da un lato all'altro) del tetto, che corrisponde all'asse dorso-ventrale (dall'alto in basso) del campo visivo. Tuttavia la mappa era invertita lungo l'asse antero-posteriore (dall'avanti all'indietro); invece di proiettare dalla parte anteriore a quella posteriore del tetto, le fibre proiettavano nell'ordine inverso. La rotazione dell'occhio allo stadio 31, o piú tardi, dava luogo a una completa inversione della mappa tetto-retinica sia lungo l'asse medio laterale sia lungo quello antero-posteriore del tetto (*si veda l'illustrazione in basso a pagina 63*). D'altro canto, la struttura dell'occhio è determinata nelle prime fasi dell'embrione, in modo che un occhio ruotato rimane sempre anatomicamente capovolto, anche quando l'operazione è stata eseguita prima dello stadio 30 e perciò dà luogo a una proiezione normale.

Questo vuol dire che c'è un cambiamento nell'embrione, durante un periodo critico che dura da 5 a 10 ore, tra gli stadi 28 e 32, che altera la risposta del sistema tetto-retinico per una rotazione dell'occhio di 180°. Le nostre ricerche piú recenti sono cominciate con una serie di esperimenti di controllo, che hanno dimostrato che la differenza tra la risposta allo stadio 28 e la risposta allo stadio 32 non consisteva semplicemente nell'inversione o non inversione delle mappe tetto-retiniche da parte di occhi trattati chirurgicamente. Abbiamo asportato degli occhi e quindi li abbiamo riimpiantati sia in una posizione normale, sia ruotati con angolo variabile tra 90° e 270°. Abbiamo fatto una mappa delle risultanti proiezioni tetto-retiniche dopo che la rana aveva subito la metamorfosi da girino ad adulto. La mappa tetto-retinica era sempre normale, quale che fosse l'angolo di rotazione dell'occhio, quando l'operazione era stata eseguita allo stadio 28 o piú precocemente. Se l'operazione veniva eseguita allo stadio 30 o dopo, la mappa tetto-retinica era sempre ruotata di un angolo esattamente uguale a quello di rotazione dell'occhio.

Questi esperimenti rendono possibili tre importanti conclusioni. Per prima cosa, l'insieme delle specificità di luogo che si sviluppa in occhi che sono stati ruotati prima del periodo critico è organizzato spazialmente in accordo con le posizioni successive all'operazione delle cellule retiniche e non tiene nessun conto delle posizioni che le cellule avevano prima dell'operazione. In occhi ruotati dopo il periodo critico, perciò, ha importanza la posizione delle cellule prima dell'operazione.

In secondo luogo, la dipendenza di posizione della specificità di luogo ha due componenti assiali, una relativa all'« indirizzo » antero-posteriore della cellula e l'altra all'« indirizzo » dorso-ventrale. I cambiamenti durante il periodo critico si verificano in due stadi, ciascun relativo a uno dei componenti assiali. Ruotando un occhio in pieno periodo critico è possibile produrre cellule gangliari con specificità di luogo ibride: la componente antero-posteriore è in relazione alla posizione dell'occhio precedente l'operazione e la componente dorso-ventrale è relativa alla posizione successiva all'operazione.

In terzo luogo, anche allo stadio 32, quando nessun assone retinico ha ancora avuto origine dall'occhio e la maggior parte delle cellule gangliari deve ancora formarsi, la rotazione dell'occhio dà luogo a una corrispondente rotazione dell'intera mappa tetto-retinica che coinvolge tutte le fibre del nervo ottico dell'occhio adulto. Sembra dunque che, nei primi stadi dell'embrione, venga stabilito un programma di sviluppo che interessa tutto l'insieme di specificità della retina dell'adulto.

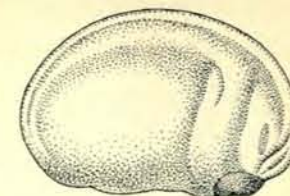
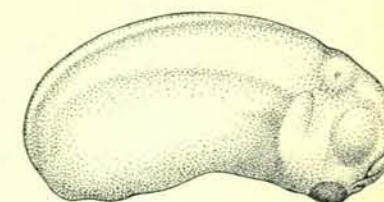
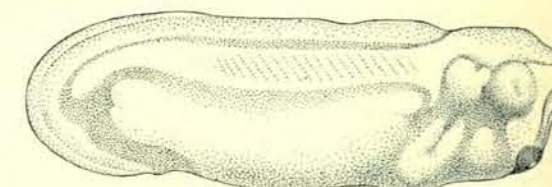
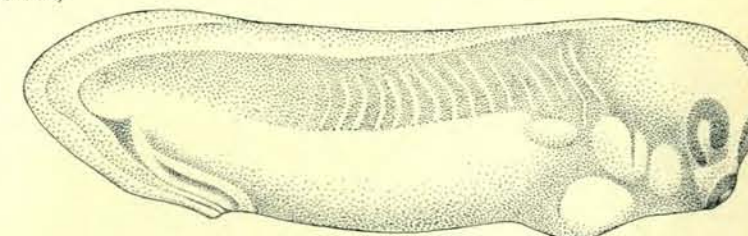
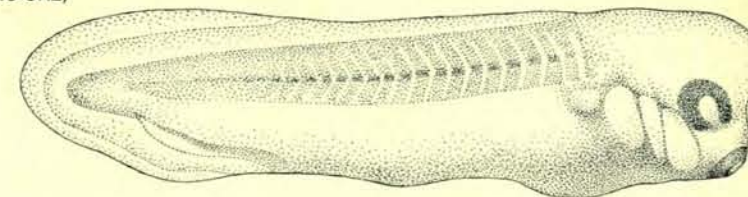
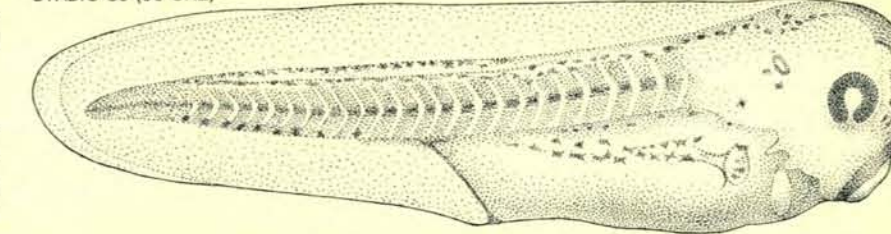
Questi esperimenti sono stati fondamentali per il nostro lavoro successivo, inoltre hanno dimostrato una qualità peculiare di questo sistema: in molte e diverse condizioni sperimentali il sistema risponde essenzialmente in un modo duplice. Cioè, se noi cambiamo la posizione anatomica dell'occhio in diverse condizioni sperimentali otteniamo o una mappa normale o una mappa ruotata.

Siamo in grado di cambiare l'orientazione dell'occhio in un trapianto tra embrioni a stadio di sviluppo diverso e di mettere in relazione i risultati con lo sviluppo dell'occhio e dell'embrione ospite. Possiamo inoltre trasferire l'occhio in una coltura di tessuto e cambiare la sua orientazione quando lo trapiantiamo nuovamente in un embrione ospite, e confrontare i risultati ottenuti con i risultati di esperimenti di trapianto diretto. Per finire, siamo in grado di privare ripetutamente l'occhio della sua orientazione con una serie di trasferimenti da un embrione donatore a uno o piú ospiti intermedi e quindi a un ultimo embrione portatore in cui si possono sviluppare le connessioni tetto-retiniche; il risultato di un simile esperimento di trapianto in serie può essere messo in relazione con le varie posizioni e orientazioni che l'occhio ha durante la sua storia.

Un esperimento siffatto di trapianto in serie è stato concepito per dimostrare che le specificità di luogo sono reali proprietà cellulari dipendenti dalla posizione e non soltanto la manifestazione di una corrispondenza temporale tra la data di nascita delle cellule della retina e l'arrivo delle fibre ottiche nel tetto. Noi ritardavamo l'arrivo delle fibre nel tetto trapiantando un occhio allo stadio 32 sul fianco di un altro embrione e lasciandolo lí per 30 giorni. In questo tempo, molti strati di cellule si aggiungevano al margine della retina e molte cellule si aggiungevano al tetto. Quando l'occhio veniva di nuovo trapiantato nella cavità, nella sua normale posizione anatomica, e messo in grado di formare connessioni col cervello, per la prima volta si aveva una mappa tetto-retinica normale (*si veda l'illustrazione in alto a pagina 64*).

Questo risultato dimostra che le molte cellule prodotte nella retina e nel tetto in un periodo di 30 giorni non trovavano alcuna difficoltà nello stabilire rapidamente connessioni ordinate. Inoltre, quando l'occhio veniva inserito nella cavità in posizione capovolta, la proiezione tetto-retinica risultante era ordinata, ma capovolta, indicando che il sistema aveva completato un programma stabilito prima dell'intervento chirurgico. Questi esperimenti dimostrano che la specificità di luogo è una caratteristica cellulare stabile e che proprio una simile proprietà permanente, dipendente dalla posizione, deve regolare lo sviluppo embrionale della mappa tetto-retinica.

Un simile esperimento poneva il seguente problema: da quali posizioni dipende la specificità di luogo? Ci deve essere chiaramente un'interazione tra l'occhio e le linee assiali dell'embrione, poiché gli assi dell'occhio sono allineati con gli assi principali dell'embrione. Il problema è se l'occhio utilizza il corpo solo come fonte di stimoli per stabilire dei propri assi allineati in modo adeguato (e quindi specifici indirizzi cellulari individuali lungo i propri assi) o se le singole cellule retiniche vengono « istruite », direttamente dall'embrione, su come disporsi sulla superficie del corpo.

Per provare queste alternative abbiamo utilizzato embrioni che non avevano ancora raggiunto il periodo critico (*si veda la figura a pagina 64 in alto*). Occhi allo stadio 28 venivano trapiantati con orientazioni varie sui fianchi di embrioni allo stesso stadio che fungevano da ospiti intermedi, lasciati lí finché il periodo critico non era

È qui schematizzato il sistema visivo della rana. I messaggi provenienti dalle singole cellule gangliari della retina dell'occhio viaggiano lungo gli assoni delle cellule, le fibre del nervo ottico, verso il tetto ottico del cervello. I nervi ottici si incrociano nel chiasma, in modo che i segmenti del campo visivo dell'occhio sinistro vanno al tetto ottico destro e i segnali del campo visivo dell'occhio destro vanno al tetto sinistro.

Sono qui illustrati sei stadi di sviluppo dell'embrione di *Xenopus*. I disegni hanno tutti la stessa scala, l'embrione è lungo circa due millimetri allo stadio 22 e sei millimetri allo stadio 39. Per ogni stadio è riportato il tempo a partire dalla fecondazione. L'embrione si distingue bene intorno allo stadio 30 e comincia a nutrirsi allo stadio 42.

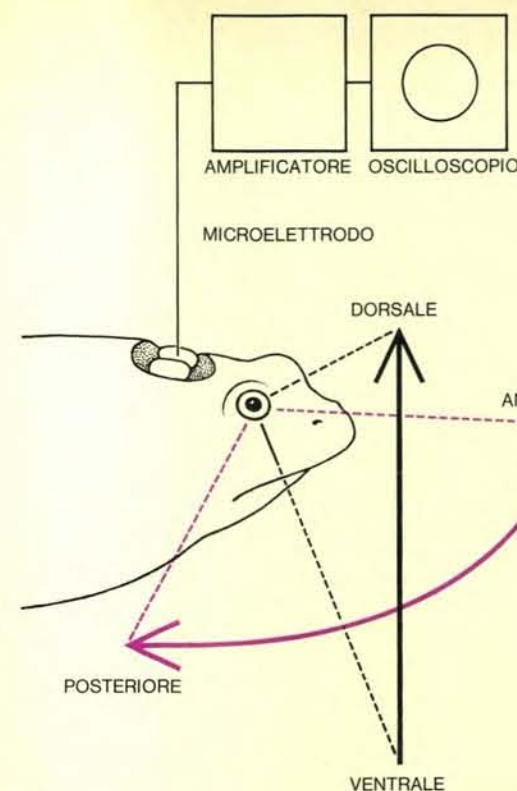
completamente trascorso e quindi trapiantati nuovamente nelle orbite degli embrioni che fungevano da portatori finali. Quando le proiezioni tetto-retiniche venivano localizzate dopo la metamorfosi, il complesso delle specificità di luogo era in ogni caso organizzato spazialmente in accordo con l'orientazione dell'occhio sul corpo dell'ospite intermedio. È chiaro perciò che le linee assiali non esistono solo per i tessuti che normalmente circondano l'occhio nella testa, ma sono anche alla portata di un occhio che si trovi sul fianco del corpo. Questi esperimenti dimostrano che le specificità di luogo dipendono dalla posizione delle cellule gangliari nell'occhio, e che devono trarre origine dalle informazioni sull'indirizzo delle cellule lungo gli assi propri dell'occhio e non da un qualsiasi valore assoluto o misura di posizione sulla superficie del corpo.

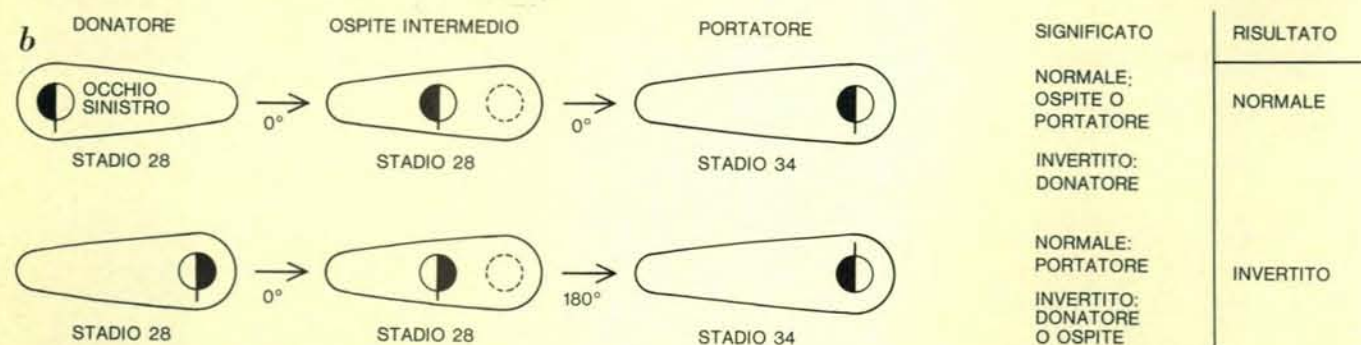
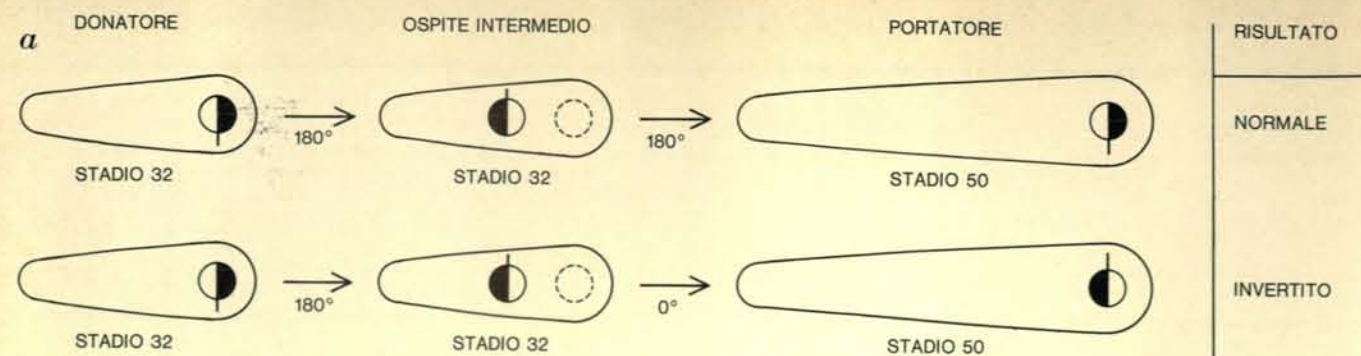
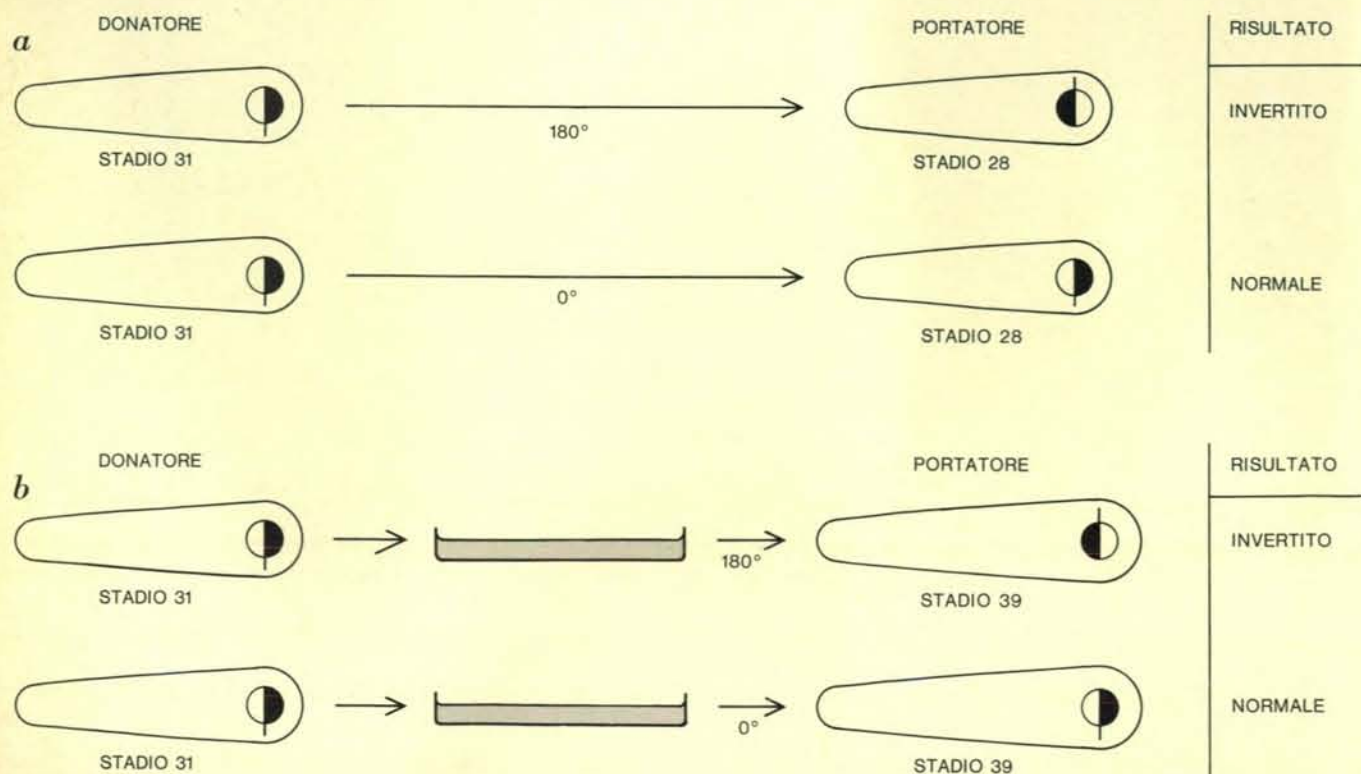
Nella descrizione di questi risultati e dei seguenti, ci siamo rifatti alle teorie di Lewis Wolpert della Middlesex Hospital Medical School di Londra, il quale ha formulato un'ipotesi sull'origine delle differenze dovute alla posizione nelle popolazioni di cellule embrionali. Nella sua notevole astrattezza la terminologia di Wolpert è compatibile col fatto che i nostri esperimenti forniscono al massimo grado delle caratterizzazioni operative; dobbiamo evitare di fare delle affermazioni non provate per quanto riguarda i meccanismi e i modi con cui le singole cellule gangliari ottengono informazioni sulla loro posizione retinica e sviluppano la corrispondente specificità di luogo. Perciò pensiamo che i cambiamenti che si verificano durante il periodo critico dallo stadio 28 al 32 sono gli eventi programmatori che legano la retina ad assi di riferimento permanenti e specificano le informazioni di posizione che tutte le cellule gangliari utilizzano per sviluppare le loro precise specificità di luogo. Noi parliamo di un processo o periodo di « specificazione », e chiamiamo gli stadi del sistema, prima e dopo il periodo critico, « non specificati » e « specificati ».

La transizione dalla condizione non specificata del sistema nello stadio 28 alla condizione specificata degli stadi 31 e 32 implica un cambiamento stabile e irreversibile della stessa popolazione delle cellule retiniche (*si veda la figura in basso a pagina 64*). In un esperimento, un occhio allo stadio 31 veniva « retro-trapiantato » nella cavità orbitale di un embrione allo stadio 28, che, come sappiamo, presenta tutte le condizioni necessarie per stabilire degli assi allineati in modo appropriato in un occhio allo stadio 28 ruotato o trapiantato. In un secondo esperimento un occhio allo stadio 31 veniva mantenuto da 6 a 10 giorni in una coltura di tessuto prima di essere nuovamente trapiantato nella cavità di un ospite allo stadio 39. In entrambi i casi, dagli occhi trapiantati di nuovo con orientazione normale, si sviluppavano le mappe tetto-retiniche normali, ma quando gli occhi venivano trapiantati in una posizione ruotata, la mappa tetto-retinica era parimenti ruotata. Questo vuol dire che né le condizioni presenti in un embrione allo stadio 28, né il completo isolamento dall'embrione erano in grado di invertire o annullare la fase di specificazione che era stata raggiunta dall'occhio allo stadio 31.

La giunzione tra due cellule retiniche si presenta come una doppia linea spessa in questa fotografia al microscopio elettronico di Marija Duda. A livello della giunzione lo spazio tra le cellule si restringe da oltre 100 angstrom a circa 30. Queste giunzioni sono visibili tra parecchie cellule retiniche allo stadio 28, ma sono meno frequenti agli stadi 29 e 30 e sono assenti tra le cellule che non si dividono dopo lo stadio 31. Le giunzioni servono apparentemente come canali per le comunicazioni intercellulari.

Questi esperimenti dimostrano che tra gli stadi 28 e 31 la retina subisce un cambiamento stabile e irreversibile che apparentemente rende la popolazione delle cellule retiniche refrattaria a informazioni relative a successivi cambiamenti di posizione. Le singole cellule subiscono questo cambiamento e manifestano questa stabilità indipendentemente, oppure il cambiamento si verifica in tutta la popolazione delle cellule retiniche e richiede l'integrità strutturale dell'occhio? Per rispondere a questa domanda abbiamo unito la metà nasale di un occhio destro allo stadio 32 con la metà temporale di un occhio sinistro allo stesso stadio per formare un occhio strutturalmente completo. Se ciascun semi-occhio embrionale avesse dato origine alle specificità della regione corrispondente di un occhio normale, le due metà dell'occhio ricostituito avrebbero presentato proiezioni verso regioni diverse del tetto. Invece ciascuna metà presentava delle proiezioni verso l'intero tetto, con punti specularmente simmetrici in ogni semi-retina che si proiettava verso lo stesso luogo nel tetto (*si veda la figura a pagina 65*). Risultati simili furono ottenuti dopo fusione di frammenti di semi-occhi e perfino dopo una semplice bisezione chirurgica dell'occhio allo stadio 32. In altre parole, il programma originale per la disposizione spaziale delle specificità di luogo non veniva espresso dopo la soppressione dell'integrità strutturale dell'occhio allo stadio 32, suggerendo che la stabilità della condizione specificata dell'occhio allo stadio 32 richiede una popolazione minima di cellule retiniche e non si estende alla singola cellula retinica; ma può richiedere l'integrità strutturale dell'intero occhio.

La transizione dalla condizione non specificata a quella specificata trae origine da meccanismi insiti nella popolazione di cellule retiniche oppure disposti altrove nell'embrione? Abbiamo trapiantato un occhio embrionale molto giovane da un donatore allo stadio 23, in posizione ruotata, in un ospite allo stadio 28, facendolo passare attraverso il periodo critico di quell'ospite e quindi trapiantandolo nuovamente in un embrione allo stadio 32 che fungeva da portatore definitivo. È importante notare che quest'ultimo trapianto veniva eseguito prima che lo stesso occhio avesse raggiunto il suo periodo critico. Questi occhi sviluppavano sempre delle proiezioni tetto-retiniche normali senza tener conto della loro orientazione nella cavità dell'ultimo portatore, indicando che l'insieme delle specificità di luogo era organizzato spazialmente lungo

La mappa tetto-retinica mette in relazione il campo visivo con la risposta elettrica nel tetto. Il campo è definito da due assi: il dorso-ventrale, o dall'alto in basso, e l'antero-posteriore, o dall'avanti all'indietro (*a sinistra*). Questi assi corrispondono a due assi nel tetto (*a destra*). L'asse antero-posteriore si proietta nel tetto dall'avanti all'indietro; il dorso-ventrale da un lato all'altro.

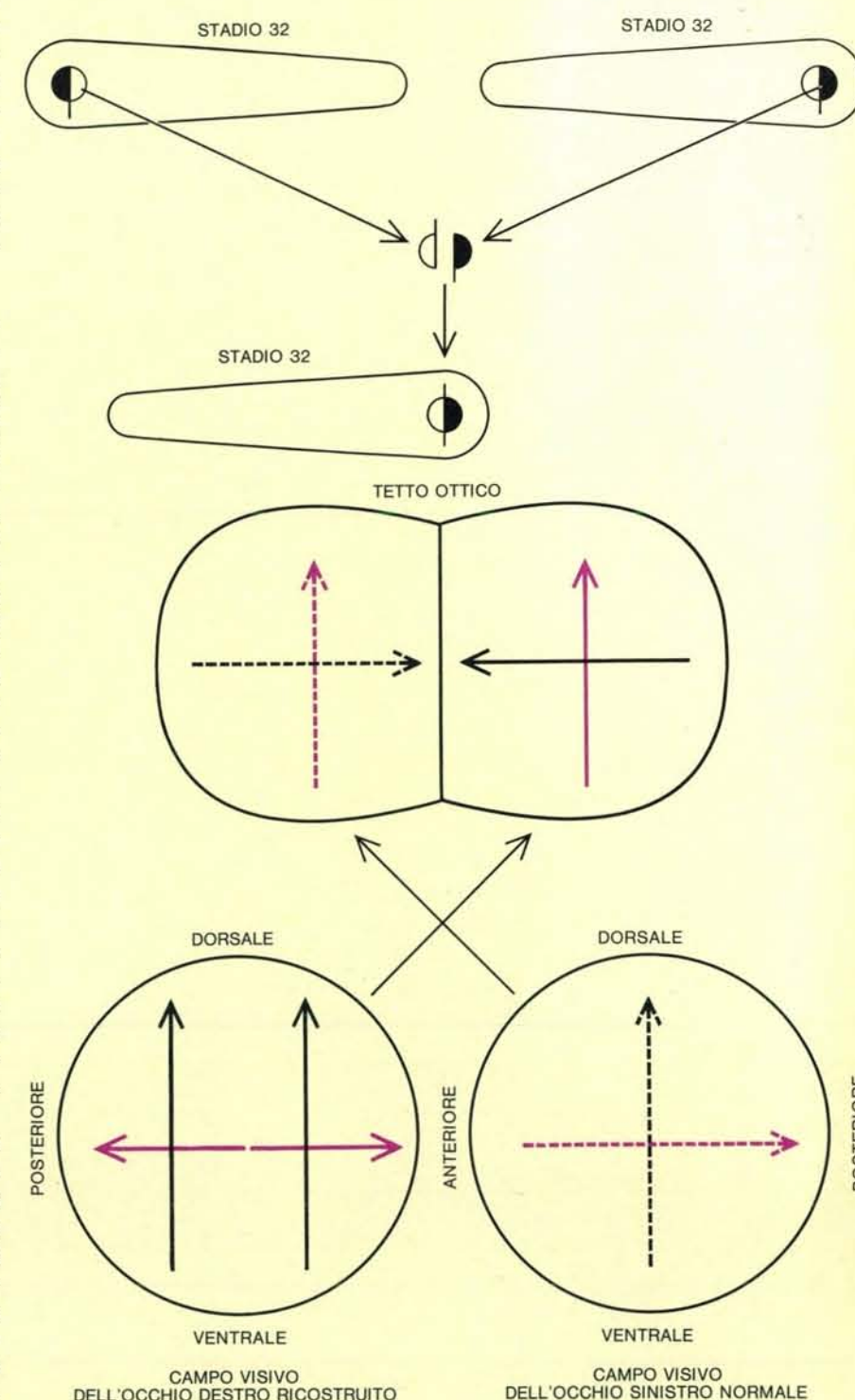
Una inversione dell'occhio a vari stadi dello sviluppo embrionale ha effetti diversi sulle proiezioni tetto-retiniche che vengono in seguito mappate nella rana. Se l'occhio destro viene ruotato di 180 gradi allo stadio 28, la mappa tetto-retinica è normale (*a sinistra*). Se l'occhio viene ruotato allo stadio 30, la mappa risulta invertita in un asse, l'antero-posteriore (*al centro*). Se l'occhio viene ruotato allo stadio 31 o piú avanti, la mappa è completamente invertita (*a destra*). I cambiamenti di proiezione del campo visivo nel tetto vengono mostrati in questi disegni schematici come cambiamenti del campo visivo e non del tetto, poiché la procedura sperimentale implica il movimento dello stimolo visivo mantenendo la posizione dell'elettrodo nel tetto.

I trapianti in serie dimostrano la stabilità (*a*) e l'origine (*b*) della informazione di posizione. Un occhio privato di connessioni con il tetto per 30 giorni e quindi nuovamente trapiantato forma una mappa ordinata che è orientata secondo la sua orientazione anatomica (*a*). Anche gli occhi che passano il periodo critico fuori dalla cavità orbitaria danno luogo a mappe ordinate (*b*). Da chi ha origine la loro specificità: dal donatore, dall'ospite o dall'ultimo portatore? Viene dato il significato della relazione tra situazione di partenza e risultato. I risultati sperimentali dimostrano come in entrambi i casi l'orientazione della mappa sia appropriata all'orientazione dell'occhio nell'ospite, vale a dire all'orientazione che l'occhio ha durante il periodo critico.

Il periodo critico tra gli stadi 28 e 32 implica un cambiamento stabile e irreversibile nelle cellule retiniche, come si dimostra con un «retro-trapianto» in uno stadio embrionale piú precoce (*a*) e con colture di tessuto in cui l'occhio è mantenuto in vitro per 6-10 giorni (*b*). In entrambi i casi l'orientazione della mappa è quella anatomica dell'occhio nell'ultimo portatore. La condizione di specificità viene raggiunta allo stadio 31 ed è irreversibile: né l'incontro con membrane allo stadio 28 (*a*) né il totale isolamento dall'embrione mediante coltura di tessuto (*b*) può annullare questa condizione, una volta raggiunta.

assi di riferimento acquisiti dall'ultimo embrione portatore durante il periodo critico proprio dell'occhio (*si veda la figura superiore a pagina 66*).

Questo risultato dimostra che l'embrione conserva la sua capacità di fornire delle linee assiali anche dopo aver passato il normale periodo di specificazione. Inoltre una specificazione precoce non è affrettata dal contatto dell'occhio con un embrione che sta attraversando il periodo critico se l'occhio stesso non ha raggiunto questo stadio. Noi pensiamo che la transizione dalla condizione non specificata sia controllata da meccanismi insiti nella popolazione di cellule retiniche, che entrano in azione quando le cellule raggiungono uno stadio predeterminato della loro differenziazione.

L'occhio possiede una qualsiasi informazione sulla sua orientazione durante la condizione non specificata, prima dello stadio 28? Occhi rudimentali di embrioni molto precoci (stadi dal 22 al 25) sono stati coltivati in vitro per periodi dai due ai quattro giorni. Alcuni di questi espianti arrivavano alla forma e all'architettura cellulare di occhi che erano giunti allo stadio 38 o 39; essi davano anche origine nella coltura a molte fibre del nervo ottico. Questi espianti particolarmente vitali e maturi venivano poi nuovamente trapiantati in embrioni portatori allo stadio 39. Quando le proiezioni tetto-retiniche venivano mappate dopo la metamorfosi, le mappe erano normali nel caso degli occhi che erano stati trapiantati nell'ultimo portatore con la orientazione anatomica normale, ma erano invertiti per gli occhi che erano stati trapiantati con una orientazione ruotata (*si veda la figura in basso a pagina 66*). Inoltre anche quando l'esperimento veniva ripetuto con ospiti allo stadio 28, che notoriamente presentano tutte le condizioni necessarie per stabilire un nuovo insieme di assi in un occhio non specificato, i risultati erano gli stessi.

Questi esperimenti dimostrano che la specificazione può verificarsi in colture di tessuto completamente isolate dall'embrione, che l'occhio già allo stadio 22 contiene un insieme di assi di riferimento allineati in modo appropriato con gli assi dell'embrione, e che la specificazione in una coltura dà luogo a un insieme di specificità di luogo che sono organizzate spazialmente intorno a questi assi precoci dell'occhio. Lo stesso processo di specificazione – la transizione durante il periodo critico dallo stadio 28 allo stadio 31 – è quindi principalmente un cambiamento da una condizione reversibile a una irreversibile, e non richiede alcuna informazione da parte dell'embrione. La condizione non specificata deve essere stabile, ma reversibile: le relazioni assiali di un occhio non specificato possono sopravvivere in una coltura fin quando l'occhio è pronto per la specificazione, ma possono essere sostituite da nuove relazioni assiali facenti seguito a un diretto trapianto dell'occhio.

Per concludere, esaminiamo i possibili cambiamenti di condizione e i meccanismi di controllo nello sviluppo di una proiezione tetto-retinica normale di un occhio ruotato allo stadio 25. Inizialmente l'occhio rudimentale è in

La stabilità della condizione di specificità è stata alterata con una operazione in cui sono stati combinati due mezzi occhi: la metà temporale di un occhio sinistro, ruotato di 180 gradi, e la metà nasale di un occhio destro con la normale orientazione. Gli occhi allo stadio 32 hanno superato il periodo critico, ma ognuna delle due metà non ha proiezioni verso punti diversi del tetto. Ciò dimostra che la specificità è una proprietà dell'occhio intero o almeno di una popolazione minima di cellule retiniche.

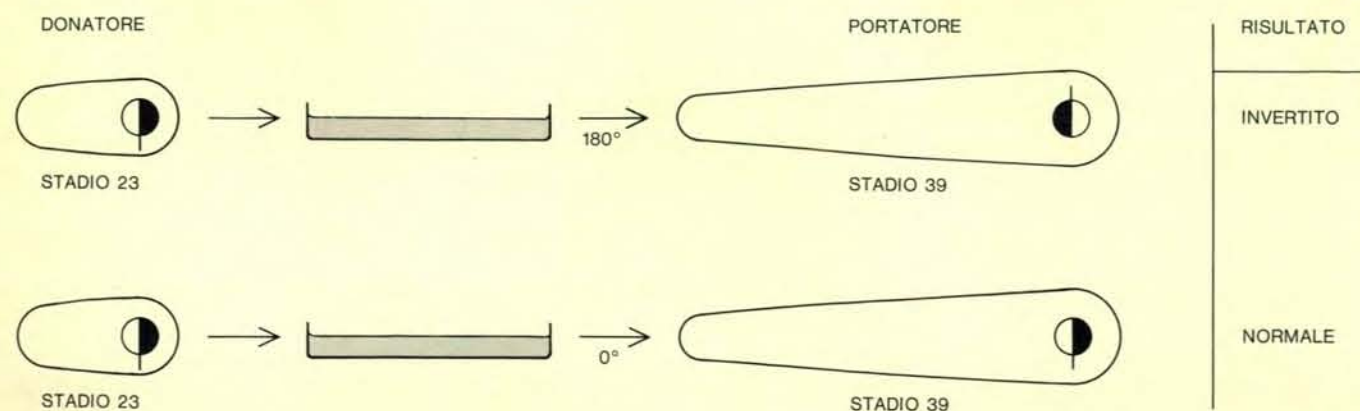
una condizione stabile, ma reversibile, possedendo un insieme primitivo di assi ad angolo retto allineati con gli assi antero-posteriore e dorso-ventrale dell'embrione. Un occhio ruotato allo stadio 25 interagisce con le linee assiali dell'embrione e la condizione iniziale si modifica in una nuova condizione che è anche stabile, ma reversibile. L'insieme di assi originale è stato rimpiazzato da un nuovo insieme di assi dell'occhio, che sono ancora allineati con gli assi principali dell'embrione.

La transizione successiva, che ha origine quando il programma di sviluppo della popolazione di cellule retiniche raggiunge un punto critico, « blocca » l'asse antero-posteriore dell'occhio e lo specifica come l'asse di riferimento per la componente antero-posteriore dell'informazione di posizione per le singole cellule gangliari. La terza condizione, poi, è caratterizzata da un asse dorso-ventrale stabile, ma reversibile e da un asse antero-posteriore stabile e irreversibile. Infine, entra in azione un altro meccanismo, anche sotto controllo retinico, che blocca l'asse dorso-ventrale dell'occhio come asse di riferimento per la componente dorso-ventrale dell'informazione di posizione per tutte le cellule gangliari. La quarta condizione è stabile e completamente irreversibile nell'occhio intatto.

Nel corso di un normale sviluppo ci può essere un meccanismo per mantenere l'occhio in sviluppo allineato anatomicamente con gli assi del corpo, rendendo cosí superfluo un continuo riallineamento degli assi retinici. (Per esempio, abbiamo notato che quando l'occhio è ruotato di 90° o meno, esso spesso ruota di nuovo all'indietro verso la sua normale orientazione anatomica.) D'altro canto, tre tipi di prove testimoniano della rapidità con cui gli assi retinici possono essere rimpiazzati, e quindi della facilità di un tale meccanismo di ricambio assiale per mantenere gli assi della retina e del corpo allineati durante il normale sviluppo.

Per prima cosa, non abbiamo mai trovato una parziale inversione dell'asse retinico per centinaia di casi di rotazione dell'occhio, e gli assi della retina e del corpo umano sono sempre perfettamente allineati dopo l'inversione di un occhio non specificato. In secondo luogo, passano meno di due ore tra lo stadio 30, quando l'inversione dell'occhio risulta in una sostituzione dell'asse retinico dorso-ventrale, e lo stadio 31 in cui l'asse dorso-ventrale è specificato irreversibilmente. Terzo, la sostituzione assiale si verifica molto rapidamente in un occhio invertito. Abbiamo ruotato un occhio per due ore allo stadio 26 e lo abbiamo trasferito in una coltura di tessuto per uno o piú giorni prima di trapiantarlo, con la normale orientazione, nell'orbita di un embrione allo stadio 39. La proiezione tetto-retinica risultante era invertita, dimostrando che gli assi retinici si erano formati durante il periodo di due ore di inversione dell'occhio prima che l'occhio venisse trasferito alla coltura di tessuto.

In qual modo le cellule gangliari che sono nate nella retina del girino « scoprono » il piano di sviluppo stabilito nei primi stadi dell'embrione? La trasmissione di informazioni di posizione alle cellule gangliari neo-formate può avvenire attraverso la linea cellulare, per comunicazione intercellulare o mediante una combinazione dei due meccanismi. Ogni cellula madre nella retina allo stadio 32 può essere destinata, come capostipite di una linea di cellule con un preordinato insieme di specificità. Tuttavia non è cosí; l'esperimento con un occhio formato da due metà di retina allo stadio 32 mostra che i frammenti non esprimono l'insieme di specificità prevedibile in base ai loro luoghi di origine retinica, ma piuttosto assumono un insieme di specificità completamente diverso. Inoltre, interferendo con la produzione di cellule al margine della retina (sia interrompendo la divisione cellulare, sia uccidendo la maggior parte delle cellule madri) non si impedisce lo sviluppo di mappe tetto-retiniche organizzate normalmente. È perciò molto probabile che le cellule gangliari neo-formate al margine della retina acquistino in qualche modo informazioni relative alla posizione dalle cellule piú vecchie al centro della retina.

L'origine della specificazione è insita nell'occhio stesso. Occhi embrionali molto giovani sono stati trapiantati successivamente su embrioni allo stadio 28 e 32. La specificazione non era affrettata dall'impianto sull'embrione allo stadio 28; ma attendeva che l'occhio giungesse al proprio periodo critico che arriva solo dopo il suo trapianto sull'ultimo portatore allo stadio 32.

L'informazione riguardante la posizione può essere specificata in occhi tenuti in coltura di tessuto. Occhi rudimentali venivano coltivati per 2-4 giorni e quindi nuovamente trapiantati su portatori allo stadio 39. I risultati dimostrarono che la specificazione si era verificata nella coltura sulla base degli assi di riferimento che si erano stabiliti negli occhi nell'embrione ai primi stadi; la specificazione nel periodo critico implica un cambiamento che si sviluppa verso una condizione irreversibile.

Il meccanismo con cui le singole cellule gangliari elaborano l'informazione di posizione deve ancora essere studiato. Tuttavia sono stati fatti alcuni progressi al fine di capire le basi cellulari del cambiamento che si verifica nella retina agli stadi compresi tra il 29 e il 31. Alcuni anni fa uno di noi (Jacobson) mostrò che questo cambiamento di stato si verifica appena le cellule al centro della retina cessano di sintetizzare DNA e di dividersi. La scoperta attirò l'attenzione sugli eventi dell'ultimo ciclo di divisione cellulare come origine del cambiamento dalla condizione non specificata a quella specificata.

Recentemente, John Dixon e John Cronly-Dillon dell'Università di Manchester hanno fatto l'interessante scoperta che alcune giunzioni intercellulari scompaiono dalle cellule al centro della retina di *Xenopus*, tra gli stadi 29 e 31, ma permangono al margine retinico. Le giunzioni sono state da allora identificate senz'altro come giunzioni distanziate che si ritiene generalmente servano come canali per il passaggio di piccole molecole tra cellule vicine (si veda l'articolo *Comunicazione intercellulare* di Werner Loewenstein, in « Le Scienze » n. 24, agosto 1970). Abbiamo cominciato a lavorare con Loewenstein e Birgt Rose dell'Università di Miami, studiando il trasferimento di piccole molecole marcate con fluorescenza tra le cellule retiniche a vari stadi di sviluppo. I nostri esperi-

Questo esperimento dimostra la completezza del complesso di specificità del terzo occhio, che viene impiantato nella stessa cavità di un occhio normale e invia le sue proiezioni verso lo stesso tetto. Dopo i passaggi che sono schematizzati (*in alto*) l'occhio operato aveva una mappa in perfetto accordo con l'occhio normale, e questo sta a indicare che doveva essere in possesso di un complesso di specificità di luogo normale.

menti hanno dato risultati incoraggianti indicanti che il trasferimento da cellula a cellula viene realmente interrotto nel periodo critico. Può essere che l'acquisizione di una informazione di posizione richieda una comunicazione intercellulare, e il distacco tra cellule può essere un evento critico nella transizione dalla condizione non specificata.

Finora abbiamo trattato solo dell'acquisizione di specificità di luogo nella retina, poiché il fenomeno ha una solida base sperimentale sulla quale è possibile fare ipotesi e concepire altri esperimenti. Prove in numero molto inferiore sono state ottenute per quanto riguarda la scelta di vie assoniche e i luoghi di connessioni sinaptiche. Un motivo per questa scarsezza di prove è che non conosciamo ancora i modi in cui le specificità di luogo retiniche vengono espresse durante lo sviluppo dei circuiti tetto-retinici.

Al fine di ovviare a questa limitazione abbiamo concepito un esperimento che dovrebbe aiutarci a stabilire quali specificità siano connesse alle cellule gangliari che compongono la mappa tetto-retinica. In questo esperimento, mettiamo un occhio normale e un occhio manipolato sperimentalmente nella stessa cavità facendo sí che essi proiettino i loro assoni verso lo stesso tetto. Le proiezioni provenienti dall'occhio normale servono da riferimento, e il grado in cui le fibre provenienti dai due occhi si mescolano nelle parti del tetto che sono comuni danno una misura della completezza dell'insieme di specificità di luogo retiniche nell'occhio operato. In tutti gli esperimenti in cui avevamo ruotato, trasferito o messo in cultura un occhio, l'occhio normale e quello operato costituivano una mappa in perfetto accordo nello stesso tetto, dividendosi tutte le regioni del tetto (*si veda la figura a fianco*). Questo dimostrava che l'occhio operato conteneva un insieme completo di specificità di luogo.

L'esperimento dei tre occhi dovrebbe dare una misura precisa dell'insieme di specificità di luogo di un frammento di occhio o di un occhio ricostruito da due o piú frammenti e permetterci cosí di scoprire i modi in cui la specificità di luogo retinica si esprime. Sapremo allora se le singole specificità di luogo possono manifestarsi indipendentemente o se possono esprimersi soltanto nel contesto dell'intero insieme di specificità presenti. Finché questi esperimenti non saranno completati, le regole che governano la crescita selettiva dei neuroni e le caratteristiche specifiche di connessione dei neuroni rimarranno un enigma.

La lingua cinese

Questa lingua melodiosa è parlata da piú persone di qualsiasi altra. Nonostante la complessità del sistema cinese di scrittura, la struttura fondamentale della lingua è molto semplice

di William S-Y. Wang

Alle persone che hanno familiarità solo con le lingue europee piú comuni, la lingua cinese appare assai diversa. Il cinese è nondimeno parlato oggi da un numero di persone piú elevato che non qualsiasi altra lingua e la letteratura cinese è la piú antica del mondo, estendendosi su un periodo di 35 secoli. Se esaminiamo la struttura della lingua cinese, ci rendiamo conto che non è eccessivamente complessa; per molti aspetti il cinese è anzi piú semplice delle lingue occidentali. Ma poiché esso differisce per alcuni aspetti fondamentali dalle lingue europee, è indispensabile avere una qualche nozione della sua struttura e del suo sviluppo storico se si vuol pervenire a una comprensione generale della natura di questo linguaggio.

Agli occhi di un occidentale il sistema di scrittura cinese appare un'assoluta novità: invece di serie di lettere alfabetiche semplici che si ripetono con grande frequenza ci sono migliaia di caratteri unici, molti dei quali hanno un aspetto incredibilmente complicato. All'orecchio la lingua ha un suono alquanto melodioso, tanto da dare l'impressione di essere quasi cantata. Quando si scava un po' sotto la superficie, le sorprese aumentano. La lingua non possiede praticamente una coniugazione per i verbi e una declinazione per i nomi. Gli inevitabili paradigmi che sono diventati un incubo per gli scolari occidentali sono del tutto assenti in una grammatica cinese. Per le varie forme del verbo « comprare », come « compro », « compri », « comprai », « comprerò » il cinese ha la forma unica *mǎi*. (Il segno sulla *a* significa che la sillaba è pronunciata con un tono che scende e poi sale. Esistono altri tre segni vocalici per indicare il tono: *á* indica per esempio un tono ascendente, *à* un tono discendente e *ā* un tono piano.) Un unico termine cinese *shū* sta per « libro », « libri », « ai libri » e « dei libri ». Per lo piú risulta abbastanza chiaro dal contesto quali siano il tempo o il modo che si devono attribuire a un verbo e quali il numero o il caso da assegnare a un nome. La lingua cinese non deve quindi preoccuparsi molto nella sua grammatica di questo tipo particolare di ridondanza. Fu forse questa semplicità strutturale della lingua a indurre l'antropologo e linguista Edward Sapir a definirla « sobriamente logica ».

La intima relazione tra la pittura e la scrittura nell'arte cinese è elegantemente esemplificata dalla figura a fronte che riproduce un dettaglio da « piante e alberi in fiore ». I caratteri sono scritti in forma tradizionale: in colonna e da destra a sinistra. Il brano di poesia, tra i rami del pino, può cosí tradursi: « La giada oscilla molle nel vento. Il rosso berillio scintilla splendente contro la neve ». La pittura, eseguita in Cina da Ch'en Shun nel 16° secolo, durante la dinastia Ming, appartiene alla collezione Avery Brundage al Centro dell'arte e cultura asiatica in San Francisco.

L'antichità del cinese

Il cinese viene spesso definito una lingua antichissima. In un certo senso quest'affermazione risulta però sviante. Tutte le lingue umane risalgono fino alle lontananze indistinte della preistoria e a tutt'oggi non siamo in grado di stabilire se possano essere ricondotte tutte a una radice comune. Quattromila anni fa gli antenati delle popolazioni cinesi parlavano una forma antica della lingua cinese, esattamente come gli antenati dei popoli di lingua inglese usavano una forma antica della lingua inglese. Poiché non si sa quasi nulla sull'origine del linguaggio nella specie umana, non siamo in grado di dire quali lingue si siano sviluppate prima e quali dopo. Nel corso della storia è accaduto piuttosto che talune lingue abbiano cambiato nome piú spesso di altre (in seguito a fatti come migrazioni o conquiste) e la novità dei nomi dà l'illusione della novità della cosa da essi indicata.

C'è però un senso in cui si può dire che la lingua cinese sia molto antica. Il sumerico è l'unica lingua a noi nota di cui esistano materiali scritti anteriori ai piú antichi documenti cinesi. La scrittura cuneiforme sumerica risale a circa 5000 anni fa, mentre i piú antichi documenti scritti cinesi che ci siano pervenuti non vanno oltre i 3500 anni fa. Il sumerico e le grafie da esso derivanti si sono però estinti molto tempo prima dell'inizio dell'Era cristiana, mentre l'ortografia cinese si è conservata fino a oggi, pur avendo ovviamente subíto importanti modificazioni stilistiche.

I documenti scritti cinesi piú antichi sono incisioni su ossa di animali e su scaglie di tartaruga; si tratta per lo piú di oracoli, che si riferiscono a eventi politici o religiosi, al tempo meteorologico o alla guerra. La storia di queste iscrizioni, scoperte verso la fine dell'Ottocento in farmacie dove le ossa venivano vendute come « ossi di drago » per il loro valore medicinale, è un capitolo ricco di colore nella storia dell'archeologia e della filologia cinese. Sono state raccolte a tutt'oggi piú di 100 000 iscrizioni. Benché il numero complessivo dei caratteri incisi su questi documenti superi il milione, il numero di caratteri diversi è piccolo. I testi degli oracoli trattavano una gamma di argomenti molto limitata e si ha perciò una continua ripetizione degli stessi caratteri. Dei 2000-3000 caratteri che compaiono sulle ossa, per lo piú scapole di bovini, e sui gusci di tartaruga, siamo oggi in grado di leggerne circa la metà.

I caratteri cinesi ci sono stati con-

	PERCENTUALE DI POPOLAZIONE
MANDARINO	68
WU	8
XIANG	5
YUE (CANTONESE)	5
KEJIA	4
GAN	2
MIN SETT.	1
MIN MERIDIONALE	2
NON CINESI	5

Carta della Cina in cui è illustrata la distribuzione dei dialetti principali della lingua cinese. Piú di due terzi della popolazione cinese parla un dialetto mandarino, la cui forma standard è rappresentata dalla parlata pechinese. Maggiore diffusione nel mondo, a opera di emigranti cinesi, hanno avuto però i dialetti della costa meridionale. All'interno della Cina esistono anche ceppi linguistici non cinesi. Le regioni a nord e a ovest sono dominate da lingue non cinesi come il mongolico e il tibetano. (I nomi delle città sono scritti generalmente secondo il sistema di romanizzazione adottato dalla Repubblica Popolare.)

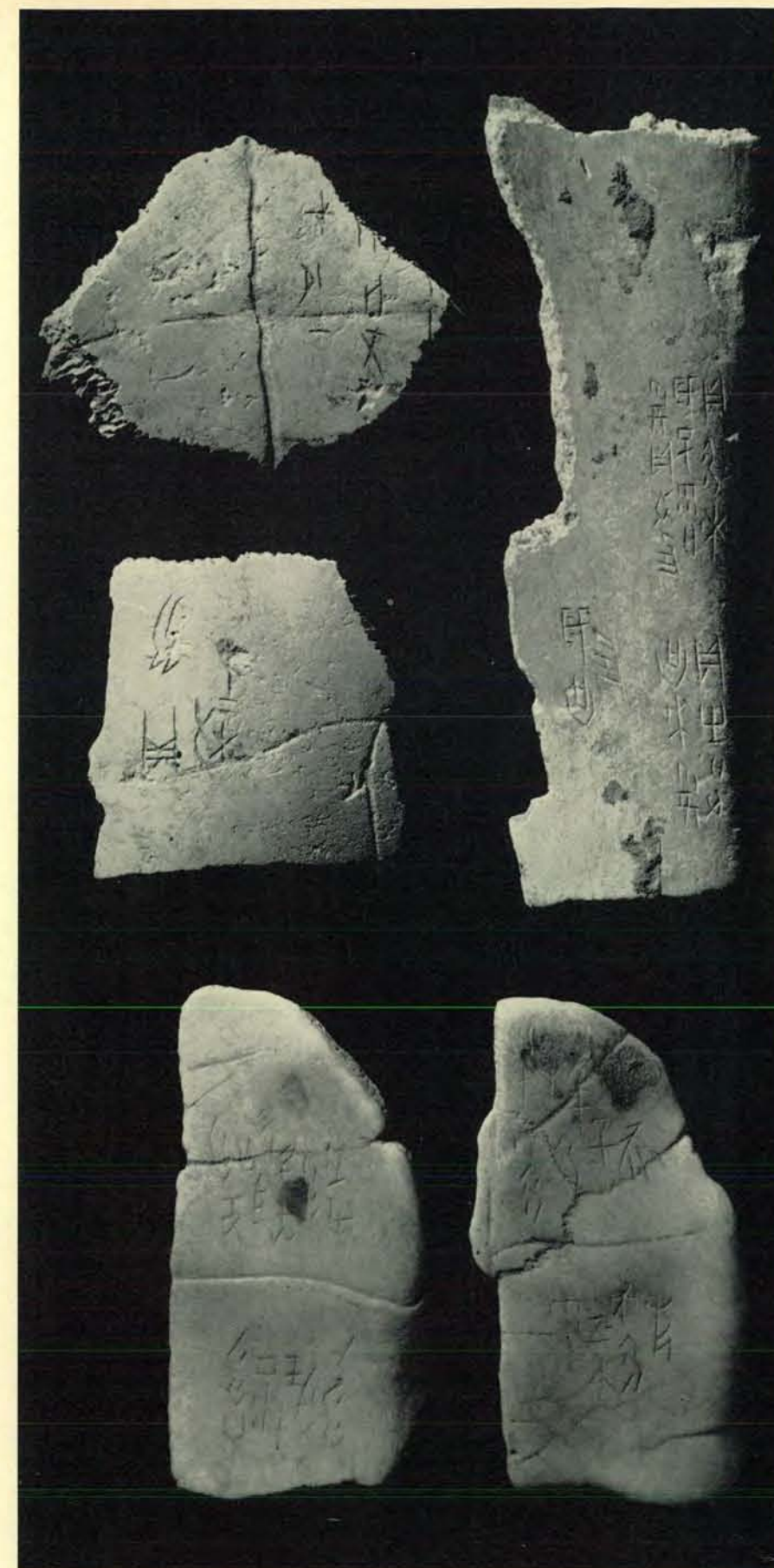
servati nel corso dei secoli dai materiali piú svariati: metallo, ceramica, rulli di pietra, giade, monete, specchi metallici, mattoni e tegole. La linea centrale nello sviluppo della scrittura cinese è stata però contrassegnata dall'uso del pennello su seta, su bambú, su legno e infine su carta. Un pennello può produrre variazioni di spessore negate a uno stilo. Tali variazioni danno allo scrivente una molto maggiore libertà artistica nella resa dei caratteri.

Alcuni tra i caratteri scritti cinesi piú antichi sono pittografici. Il carattere per « pioggia » era costituito da varie colonne di linee spezzate, e quello per « cavallo » aveva l'aspetto di un cavallo, completo di criniera e di quattro gambe (*si veda l'illustrazione a pagina 75*). Le pittografie sono però solo una minima parte dell'intero vocabolario cinese. La maggior parte delle parole non possono essere suggerite da una semplice figura.

La calligrafia, l'elegante disegno dei caratteri, è una forma d'arte che ha raggiunto in Cina vertici molto elevati e che vi è stata per molto tempo apprezzata quanto la pittura nel mondo occidentale. Per i cinesi esiste una stretta relazione tra pittura e calligrafia. Sono tipici i rotoli di seta ricoperti da un'immagine e da poche linee di caratteri, in modo che i due elementi risultino accuratamente equilibrati. A causa del loro valore artistico e della loro lunga storia, i caratteri cinesi presentano in dimensioni e forma una gamma di variabilità maggiore di quella dei caratteri di qualsiasi altro sistema di scrittura.

Il sistema di scrittura

Il sistema di scrittura cinese subí importanti mutamenti nel 1956, quando il governo della Repubblica Popolare decise di semplificare i caratteri e di adottare un sistema di romanizzazione, ossia di traslitterazione delle parole cinesi in lettere latine. I due provvedimenti avevano lo scopo di facilitare l'apprendimento della lettura e della scrittura del cinese, un passo determinante per l'unità linguistica e per l'elevazione del livello di cultura.

Per comprendere la natura dei caratteri cinesi e la loro semplificazione, dobbiamo esaminarne innanzitutto la struttura interna. Ogni carattere è costituito da due tipi di unità minori dette tratto e radicale. Grosso modo si può dire che il tratto è una linea, retta o curva, che si può considerare completata ogni volta che la penna o il pennello si solleva dalla carta. Per esempio, il carattere per « sole », che

Le iscrizioni oracolari su osso, risalenti al 1300 a.C., sono tra gli esempi piú antichi che si conoscano di scrittura cinese. Esse venivano incise su scapole di bovini e su scaglie di tartaruga ed erano usate per diffondere oracoli. La fotografia presenta alcuni esemplari conservati al Museo archeologico per l'Estremo Oriente a Stoccolma.

si pronuncia *rì*, ha il seguente aspetto:

日

Esso è costituito da quattro tratti (*si veda l'illustrazione in alto a pagina 75*), dei quali hanno importanza sia l'ordine sia la posizione geometrica. Nella lingua cinese ci sono una ventina di tratti distinti, i quali rappresentano cosí la piú stretta controparte delle 26 lettere dell'alfabeto latino. Non esiste invece alcun corrispettivo dei radicali nell'ortografia di altre lingue. La serie tradizionale consta di 214 radicali, che compaiono in quasi tutti i dizionari cinesi. Un dizionario edito nel 1971 ha fuso però alcuni radicali, riducendone il numero a 189. La maggior parte dei radicali sono nello stesso tempo anche caratteri; ad esempio il radicale « sole » e il carattere « sole » sono identici. Normalmente vengono usati migliaia di caratteri, la maggioranza dei quali non sono radicali. Ogni carattere contiene solo un radicale, solo o accompagnato da altri segni.

Il carattere per « stella », che viene pronunciato *xīng*, si scrive:

星

Esiste un carattere letterario riferito al sole del mattino (pronuncia: *lóng*), che si scrive:

昽

Possiamo vedere che entrambi i caratteri contengono il radicale « sole »:

日

Il carattere « stella » ha il radicale in alto; esso si compone anche di un altro segno, che viene pronunciato *shēng* e che ha il seguente aspetto:

生

Il carattere « sole del mattino » ha il radicale « sole » a sinistra; l'altro segno che compone la parola si pronuncia *lóng* e si scrive:

龙

Quasi tutti i dizionari e le opere di consultazione cinesi recenti ordinano i caratteri per radicali. Consideriamo per esempio il carattere per « stella ». Per trovarlo bisogna cercare prima la sezione contrassegnata dal radicale « sole »:

日

In questa sezione tutti i caratteri che contengono il radicale sole sono ordinati in base al numero di tratti contenuti nel segno residuo, che è:

生

Esso è in questo caso relativamente poco complesso, comprendendo solo cinque tratti. Dovremmo pertanto aspettarci di trovarlo alla metà circa della sezione.

Un procedimento chiave nella costruzione di caratteri può essere illustrato ricorrendo all'esempio di « stella » e « sole del mattino »:

星 昽

Entrambi i significati sono chiaramente riferiti a quello del radicale « sole ». In questi caratteri il radicale viene chiamato « classifica » o « chiave », mentre la parte restante è la « fonetica ». Le fonetiche indicano come si debbano pronunciare i caratteri. In « sole del mattino » la fonetica è *lóng*.

龙

Essa dà la sua pronuncia al carattere. Esistono d'altra parte caratteri che non sono pronunciati come imporrebbe la loro fonetica, spesso a causa di mutamenti storici del suono. Nel carattere « stella » la fonetica è:

生

che si pronuncia *shēng*, mentre la pronuncia per « stella » è *xīng*.

Per offrire un altro esempio di fonetica possiamo considerare il carattere « cavallo ». Esso si pronuncia *mǎ* e si scrive:

马

Quando la fonetica per « cavallo » viene combinata alla « classifica » per « donna », abbiamo *mā*, che significa « madre »:

妈

Quando la fonetica « cavallo » viene combinata con la classifica per « giada » abbiamo *mǎ*, che significa « agata »:

玛

Combinata con la classifica per « insetto », dà il significato « formica », e la pronuncia è ancora *mǎ*:

蚂

Quando si hanno due classifiche « bocca » sospese sulla fonetica « cavallo », il significato diventa « sgridare » e la pronuncia è *mà*:

骂

Molti caratteri sono costruiti sullo schema fonetica-classifica; cosí alla base di molti caratteri cinesi c'è un principio fonetico. Il cinese medio può spesso pronunciare correttamente un carattere che non ha mai visto prima, limitandosi a considerare attentamente la fonetica. Esaminiamo per esempio il carattere

铑

La parte radicale significa « oro ». La parte formale (fonetica) si pronuncia *lóng* e si scrive:

龙

Il carattere dovrebbe indicare con ogni verosimiglianza un metallo o un composto metallico e dovrebbe essere pronunciato come *lóng*, anche se nella lingua cinese tale carattere non esiste.

Un altro carattere inesistente mi fu inviato poco tempo fa da un amico come enigma. Il carattere si compone di tre parti:

女 上 下

La prima componente significa « donna ». Anche la seconda e la terza componente sono caratteri indipendenti; essi vengono pronunciati *shàng* e *xià* e significano rispettivamente « su » e « giú ». Assegnammo la funzione di classifica alla componente « donna », e il carattere venne ad avere questo aspetto:

娤

Per il carattere complessivo adottammo il significato « ragazza dell'ascensore ». Non c'è però un modo intuitivo di pronunciare il carattere, dato che le ultime due componenti non costituiscono una fonetica. La soluzione consistente nel pronunciarlo con due sillabe, *shàng-xià*, víola una regola generale dell'ortografia cinese, secondo la quale per un carattere c'è una sola sillaba. Benché sia stata talvolta proposta da riformatori della lingua l'introduzione di caratteri polisillabici, nessuna riforma del genere ha mai avuto successo.

Dizionari e rimari cinesi possono elencare decine di migliaia di caratteri, ma la conoscenza di 4000-7000 caratteri è sufficiente, diciamo, a poter leggere un giornale. La forma dei caratteri è stata elaborata in modo non sistematico nel corso dei secoli e molti di essi risultano assai complessi, richiedendo anche 30 o piú tratti. Il progetto di semplificazione adottato dal governo cinese ha compiuto grandi passi verso la standardizzazione della forma dei caratteri e la riduzione del numero medio di tratti che figurano in ogni carattere. Il guadagno è stato vistosissimo. Un testo scritto in caratteri semplificati può contenere un numero di tratti inferiore alla metà di quelli che lo stesso testo avrebbe richiesto prima del 1956. Questa semplificazione facilita in misura considerevole l'apprendimento della lingua scritta. Una media di cinque o sei tratti per ogni carattere non si differenzia in modo significativo dalla media di cinque o sei lettere per parola valida per gran parte delle lingue europee.

Poiché la funzione primaria dei caratteri non è quella di rappresentare suoni, la lingua scritta cinese ha conservato una grande autonomia rispetto all'evoluzione della lingua parlata. Quest'autonomia le ha consentito di garantire una continuità letteraria per migliaia di anni e di servire come mezzo di coesione per le diverse culture della Cina.

L'evoluzione della lingua

L'evoluzione del cinese parlato, come l'evoluzione di tutte le altre lingue viventi, è stata costante. Perciò molte tra le meravigliose poesie della dinastia T'ang (grafia ufficiale moderna: Táng), dal VII al X secolo, non hanno piú la rima. Se Confucio, che visse nel V secolo a.C., potesse tenere una lezione oggi in qualsiasi località cinese, la sua lingua risulterebbe incomprensibile. All'interno del territorio immenso della Cina i dialetti hanno avuto un'evoluzione cosí divergente che un abitante di Pechino non potrebbe esser certo di riuscire a ordinare un pasto in un ristorante di Canton. Rispetto ai mutamenti intervenuti nei suoni, i caratteri scritti hanno manifestato una grande costanza. La maggior parte dei caratteri usati da Confucio compaiono ancor oggi nei libri e molti di essi hanno ancora il loro significato originario. La scrittura di Confucio è piú intelligibile a un cinese moderno che non, per esempio, una pagina del poema epico anglosassone dell'VIII secolo, *Beowulf*, a un americano. Analogamente, benché le poesie T'ang non conservino la rima, sono ancora molto apprezzate in tutta la Cina grazie al fatto che il loro messaggio visivo si conserva inal-

丨 冂 月 日

Il carattere per « Sole » è formato con quattro tratti di pennello: prima un tratto verticale, poi un tratto ad angolo, quindi il tratto interno e infine la chiusura. Sono importanti sia la successione dell'operazione sia la posizione geometrica dei tratti. Tutti i componenti di un carattere dovrebbero risultare grosso modo inscritti in un quadrato.

	VECCHIO	SEMPLIFICATO
SOLE (rì)	日	日
STELLA (xīng)	星	星
SOLE DEL MATTINO (lóng)	曨	昽
CAVALLO (mǎ)	馬	马
MADRE (mā)	媽	妈
AGATA (mǎ)	瑪	玛
FORMICA (mǎ)	螞	蚂
RIMPROVERARE (mà)	罵	骂

Nel 1956 la Repubblica Popolare Cinese introdusse caratteri semplificati. I caratteri semplici, come i primi due, furono conservati inalterati. I restanti caratteri, alquanto piú complessi, furono semplificati attraverso l'eliminazione di sei tratti ciascuno.

terato. Lette ad alta voce a Pechino, a Shanghai o a Canton, le poesie hanno un suono completamente diverso venendo lette nei diversi dialetti. È come se la somma « 6 + 7 = 13 » fosse letta a voce alta in inglese, in tedesco o in norvegese. Anche in Giappone un cinese che non abbia una conoscenza neppure sommaria del giapponese può riuscire a comunicare abbastanza bene mediante la scrittura. I caratteri cinesi sono un mezzo di comunicazione efficace anche nella Corea e nel Vietnam. L'indipendenza dei caratteri dalla lingua parlata ha consentito loro di fungere da nucleo di cultura in gran parte dell'Asia Orientale per molti secoli.

La seconda importante decisione in campo linguistico presa dal governo cinese fu quella di adottare un sistema di scrittura fondato sull'alfabeto latino. Questo sistema è chiamato Pinyin, che significa letteralmente « suono scritto ». Tutte le parole cinesi scritte in lettere in questo articolo sono in Pinyin. Il governo ha sottolineato con cura che il Pinyin non intende sostituire i caratteri, ma fornire un aiuto nell'apprendimento della pronuncia. L'interruzione dell'uso dei caratteri cinesi priverebbe le future generazioni della Cina di una ricca e importante eredità culturale.

I toni sono usati per modificare il significato di parole cinesi. Il cinese standard possiede solo quattro toni: discendente (come in *mà*), ascendente (*má*), piano (*mā*) e cadente o tronco (prima discendente e poi ascendente, *mǎ*). Le tracce dell'oscillografo a destra mostrano la frequenza fondamentale della voce dell'autore nel pronunciare le parole. Nelle lingue europee la variazione di tono è usata per esprimere stati d'animo diversi e il significato delle parole non cambia. In cinese il mutamento di tono ha sul significato di una parola lo stesso effetto che può avere in una lingua europea il mutamento di una vocale (« moda », « modo ») o di una consonante (« moda », « coda »).

Il cinese standard

Per discutere il sistema di scrittura Pinyin dobbiamo prima esaminare il sistema di suoni del cosiddetto cinese standard, che si fonda sul dialetto di Pechino ed è ora insegnato in tutte le parti della Cina. È un sistema molto chiaro, per vari aspetti piú semplice delle facili sillabe dello spagnolo. Soltanto un aspetto risulta ostico agli europei: i toni. La qualità musicale della lingua cinese parlata dipende dal fatto che quasi tutte le sillabe devono avere uno dei quattro toni (o modulazioni della voce) fondamentali. Questi toni sono indicati da segni diacritici sovrapposti alle vocali. Il fenomeno dei toni pare essere limitato al cinese e ad alcune lingue dell'Asia Sud-orientale che hanno subíto fortemente l'influenza del cinese.

Uno dei piú antichi accenni ai toni che si incontrano nella letteratura risale al VI secolo. Quando l'imperatore di Liáng chiese a uno dei suoi dotti, Zhōu Shě, che cosa si intendesse con i quattro toni, Zhōu rispose con un'elegante illustrazione:

天子圣哲

Essa significa « il figlio del cielo è divino e saggio ». (Gli imperatori cinesi erano tradizionalmente considerati figli del cielo e divini, fossero o no saggi.) La bellezza della risposta di Zhōu risiede nel fatto che la prima parola della frase illustra il primo tono nella pronuncia del tempo, la seconda parola il secondo tono e cosí via.

Sostanzialmente ogni sillaba cinese ha un'intonazione particolare. Un cambiamento di tono altera il significato esattamente come accade in una lingua europea quando si cambia una consonante o una vocale. In italiano si usa un tono ascendente per « nó? » e un tono discendente per « nò! ». I diversi toni comportano una diversa altezza, ma il significato della parola rimane lo stesso. In cinese, invece, *má*, con un tono ascendente, significa « canapa » e *mà*, con un tono discendente, significa « rimproverare ». I due significati non sono piú legati tra loro di quanto sarebbero se noi dovessimo mutare la vocale per avere *mì*, che significa « miele », o se dovessimo cambiare la consonante per avere *pà*, che significa « temere ».

Il cinese standard ha un totale di quattro toni: ascendente, discendente, piano e cadente (*si veda l'illustrazione a sinistra*). Oltre al tono, ogni sillaba deve avere anche un nucleo destinato a portare il tono: di solito si tratta di una vocale. Il tono e il nucleo sono le due componenti obbligate della sillaba cinese. Esistono anche tre componenti facoltative di una sillaba: la componente iniziale, che solitamente è una consonante; la componente media, che è un suono di passaggio; e la componente finale, che può essere o un suono di passaggio o una consonante appartenente a una classe ristretta. Una sillaba può assumere otto forme diverse.

	PINYIN	TRASCR. FONETICA	INIZIALE, CONSONANTE	FINALE		
				MEDIA, PASSAGGIO	NUCLEO, VOCALE	FINALE, VOCALE O CONSONANTE
FACCIA	liăn	ljɛ̆n	●	●	●	●
FERRO	tiĕ	tʰjɛ̆	●	●	●	
CRESCERE	zhăng	tʂăŋ	●		●	●
CAVALLO	mă	mă	●		●	
OCCHIO	yăn	jɛ̆n		●	●	●
LUNA	yuè	yɛ̀		●	●	
NASCOSTO	yĭn	ĭn			●	●
SEDIA	yĭ	ĭ			●	

Le componenti obbligate di una sillaba cinese sono il tono e il nucleo (di solito una vocale), che ne è il veicolo. Ci sono poi tre componenti facoltative: l'iniziale, che di solito è una consonante; la componente media, che è un suono di passaggio, e la finale, che può essere un suono di passaggio o una consonante. Complessivamente, una sillaba cinese può assumere otto forme.

Un carattere sorprendente delle parole cinesi rispetto alla maggior parte delle parole europee è la mancanza di gruppi di consonanti prima e dopo la vocale che costituisce il nucleo. Quando parole occidentali nelle quali figurano gruppi di consonanti sono rappresentate nel cinese, vengono tipicamente scomposte in modo tale che ogni consonante viene ad essere una sillaba a sé. « Marx » è reso per convenzione:

马克思

Esso viene pronunciato *mă-kè-sī*. Il primo carattere è quello di « cavallo », che è anche un soprannome molto diffuso in Cina.

Benché in tutte le parti della Cina siano insegnati il sistema Pinyin e il cinese standard, si dà molta considerazione alle lingue delle minoranze etniche. Secondo un libro sulle nazionalità in Cina, edito a Pechino nel 1961, circa 30 milioni di persone appartengono a gruppi minoritari. Le minoranze occupano pressappoco un terzo dell'estensione del paese e vivono prevalentemente nelle regioni occidentali e nord-occidentali. Alcuni gruppi sono molto consistenti: gli Zhuàng sono circa otto milioni e gli Uighuri circa quattro milioni. La lingua di alcuni di questi gruppi non è geneticamente legata al cinese, ma appartiene ad altre famiglie linguistiche, come l'altaica e l'austroasiatica.

Nel 1956 una relazione dell'Accademia cinese delle scienze stimò che il numero totale delle persone in Cina che parlassero un qualsiasi dialetto cinese era superiore a 500 milioni. Dei dialetti, quello di gran lunga piú diffuso era il mandarino, che era parlato da piú di due terzi del numero complessivo (*si veda l'illustrazione a pagina 72*). Il nome occidentale « mandarino » deriva dal fatto che il dialetto era la lingua dei funzionari governativi o mandarini. A esso corrisponde il termine cinese *guānhuà*, che significa « modo di parlare ufficiale ». Il cinese standard è una varietà convenzionalizzata di questo dialetto ed è noto come *guóyŭ* o « lingua nazionale ». Esso viene chiamato anche *běifānghuà*, che significa « modo di parlare settentrionale », o *pŭtōnghuà*, « modo di parlare comune ».

Fuori della Cina il cinese standard è molto meno importante. I dialetti dominanti tra i 15 milioni di cinesi nell'Asia sud-orientale sono lo yue e il min meridionale, poiché in quell'area migrarono popolazioni delle province costiere meridionali. I cinesi arrivarono negli Stati Uniti relativamente tardi, ossia intorno al 1850, quando varcarono per la prima volta l'oceano Pacifico attratti dalla scoperta dell'oro in California. Piú tardi questi immigrati costituirono il grosso della manodopera usata nella costruzione delle ferrovie negli stati dell'ovest. Gli antenati della maggior parte delle varie centinaia di migliaia di cinesi che vivono oggi negli Stati Uniti provenivano da un piccolo gruppo di villaggi raggruppati attorno a Canton in un raggio di 150 km. Le parlate di questi villaggi appartengono nettamente al gruppo dialettale yue, ma presentano differenze notevoli l'una dall'altra. Possiamo esser certi che negli ultimi cent'anni si sono prodotte differenze significative tra la parlata dei cinesi americani e quella delle popolazioni che vivono oggi nei loro villaggi d'origine attorno a Canton.

La formazione delle parole

Ogni lingua ha un patrimonio di varie migliaia di morfemi, i quali possono essere definiti i veicoli del contenuto semantico e grammaticale di base. Per esempio, la parola « apriscatole » comprende tre morfemi: « apri », « scatol- » e « e ». Alcuni morfemi hanno piú di una sillaba, come « scatol- », mentre altri, in talune lingue europee, si riducono a semplici consonanti, come la « s » del plurale dei nomi di molte lingue e dialetti europei o la « t » del participio passato in inglese. Nel cinese la maggior parte dei morfemi è costituita esattamente da una sillaba. La divisione usuale dei morfemi in tre categorie principali – nome, verbo e aggettivo – si adatta abbastanza bene al cinese. Cosí in cinese *shū* (« libro ») è un nome, *măi* (« comprare ») è un verbo e *guì* (« costoso, caro ») è un aggettivo. Questi morfemi sono caratteriz-

zati dall'avere un contenuto semantico indipendente e sono detti di solito radicali.

C'è anche una classe di morfemi di tipo funzionale che vengono applicati ai radicali per modificarne in modo sistematico il significato e per indicare le relazioni che i radicali hanno l'uno rispetto all'altro. Per esempio, morfemi funzionali aggiunti al radicale « prova » ne modificano il significato: « prove », « provato », « provando », « riprovando », ecc. La coniugazione dei verbi e la declinazione dei nomi e a volte degli aggettivi, dei pronomi ecc., pur essendo importanti in molte lingue europee, sono trascurabili nel cinese, dove peraltro ha importanza abbastanza grande il fenomeno della derivazione di una parola dall'altra, come è il caso di nomi da verbi, attraverso la variazione di tono. Per esempio, dal verbo *shǔ*, che significa « contare », deriva il nome *shù*, che significa « numero »; dal verbo *lián*, che significa « collegare », deriva *liàn*, che significa « catena »; *mó* significa « macinare » e *mò* è una « macina ». Per derivare un nome da un verbo è sufficiente in questi casi dare alla sillaba il tono discendente. Un esempio simile, in italiano, consiste nel togliere al verbo la desinenza verbale e sostituirla con la desinenza del nome: « prov-are » – « prov-a », « govern-are » – « govern-o », « url-are » – « url-o » e così via.

L'espediente della derivazione mediante il mutamento di tono non è più un fenomeno importante nel cinese standard, mentre si ritiene abbia avuto una funzione di primo piano in stadi di sviluppo anteriori della lingua cinese. Nel dialetto di Canton questo fenomeno è invece ancora molto attivo. Gli abitanti di Canton usano il mutamento di tono per formare diminutivi: « dolciumi » da « zucchero », « figlia » da « femmina » e così via.

Un espediente comunemente usato in Cina nella derivazione di una parola dall'altra è la ripetizione. Applicata ai nomi, dà loro il significato di « ogni », « tutti ». *Rén* significa « persona »,*rénren* significa « ogni persona »; *tiān* significa « giorno » e *tiāntian* « ogni giorno ». Applicata ai verbi, dà un carattere transitorio all'azione: *kàn* significa « guardare », mentre *kànkan* vuol dire « dare un'occhiata »; *zǒu* significa « camminare » e *zǒuzou* « fare una passeggiata ». Gli avverbi possono essere derivati da aggettivi mediante ripetizione e aggiunta di un suffisso *de*. Così *kuài* significa « rapido » e *kuàikuàide* « rapidamente », *lǎn* significa « pigro » e *lǎnlǎnde* « pigramente ».

Gli aggettivi bisillabi si raddoppiano in modo diverso. Mentre un verbo bisillabo, per esempio *tǎolùn* (« discutere ») si raddoppia nella forma *tǎolùntǎolùn* (« discutere un po' »), un aggettivo, per esempio, *gāoxìng* (« felice ») diventa *gāogāoxìngxìngde* (« felicemente »). Un verbo raddoppia l'intera parola, mentre negli aggettivi si raddoppiano successivamente le sillabe componenti.

Un altro modo di formare parole in cinese è attraverso la congiunzione di opposti. « Comprare » e « vendere » si combinano per formare la parola « affari », *mǎimài*. « Lungo » e « corto » si combinano per formare la parola « lunghezza », *chángduǎn*. Il significato derivato non è sempre chiaro a prima vista. Per esempio, *fǎn* significa « capovolto » e *zhèng* significa « diritto ». La parola composta *fǎnzhèng* significa « in ogni caso ».

Il numerativo è un carattere linguistico peculiare del cinese e delle lingue dei paesi vicini. Articoli, numerali e altre parti secondarie del discorso non possono precedere direttamente i nomi ad esse associati; in mezzo dev'esserci sempre un numerativo, che di solito ha un contenuto semantico trascurabile. In cinese in generale non si può dire *sān shū* (« tre libri ») o *nèi māo* (« quel gatto »), ma si deve dire *sān běn shū* (« tre *esemplari di* libro ») o *nèi zhī māo* (« quell'*esemplare di* gatto »). « Esemplare di » non è una traduzione precisa dei termini *běn* e *zhī*, che si esauriscono in pratica nella loro funzione. L'uso di tali numerativi è assolutamente indispensabile in espressioni cinesi. Questo carattere si è trasferito in molte lingue commerciali e creole basate sul cinese, nelle quali « tre banane » o « quest'uomo » sono resi come « tre pezzi di banana » o « quest'individuo di uomo ».

	FORME REGOLARI	
	TIGRE	DRAGO
FIGURE ANTICHE CIRCA 2000 a.C.		
CARATTERI SU SCAGLIA E OSSO jiǎgǔwén CIRCA 1400-1200 a.C.		
GRAN SIGILLO dàzhuàn CIRCA 1100-300 a.C.		
PICCOLO SIGILLO xiǎozhuàn 221-207 a.C.		
CARATTERE DEGLI AMANUENSI lìshū CIRCA, 200 a.C.-200 d.C.	虎	龍
CARATTERI STANDARD kǎishū DAL 100 d.C. CIRCA, A OGGI	虎	龍
CARATTERI SEMPLIFICATI jianzì DAL 100 d.C. CIRCA, A OGGI		

FORME DI SCRITTURA		
TIGRE	DRAGO	
		SCRITTURA BUROCRATICA zhāngcǎo CIRCA 200 a.C.-1700 d.C.
		STILE CORSIVO xíngshū DAL 200 d.C. CIRCA, A OGGI
		CARATTERI SEMPLIFICATI liánbǐ jiǎnzì DAL 100 d.C. CIRCA, A OGGI
		SCRITTURA "MODERNA" jīncǎo DAL 300 d.C. CIRCA, A OGGI
		SCRITTURA IRREGOLARE kuángcaŏ CIRCA 600-1700 a.C.

La figura illustra lo sviluppo storico di caratteri pittografici indicanti due tra gli animali piú potenti della mitologia cinese, la tigre e il drago, in vari periodi. Le forme pittografiche piú antiche che si conoscano di questi due animali sono riprodotte, in colore. L'illustrazione è tratta da *Introduction to Chinese Corsive Script* di S.-Y. Wang.

La formazione del periodo

Il periodo elementare presenta in cinese la successione soggetto-verbo-oggetto, come in buona parte delle lingue europee. La frase *wǒmen chī jī* significa dunque, parola per parola, « noi mangiamo pollo ». Nel cinese c'è la tendenza a eliminare o il soggetto o l'oggetto. *Wǒmen chī* (« noi mangiamo ») o *chī jī* (« mangiamo pollo ») sono pertanto due tipi di frase entrambi molto comuni. Per richiamare l'attenzione sull'oggetto, il cinese lo sposterà all'inizio della frase. *Jī wǒmen chī*, per esempio, significa ancora « noi mangiamo pollo » ma la frase è usata in questa forma per sottolineare che mangiamo pollo e non qualcos'altro. Supponiamo di applicare alla frase *wǒmen chī jī* sia l'eliminazione del soggetto sia il trasferimento dell'oggetto alla posizione iniziale. Procediamo dapprima allo spostamento dell'oggetto e otteniamo *jī wǒmen chī*; poi cancelliamo il soggetto e avremo pertanto *jī chī*. La frase risultante significherà qualcosa di simile a « pollo *A* mangia », dove l'identità di *A* risulterà normalmente chiara dal contesto. Ma *jī chī* significa ovviamente anche « il pollo mangia », dove *jī* viene a essere il soggetto del verbo. In altri termini, *jī chī* è una frase ambigua, e l'ambiguità deriva dallo spostamento dell'oggetto in posizione iniziale e dalla soppressione del soggetto.

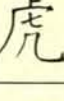
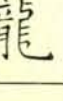
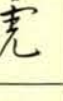
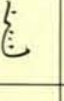
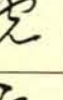
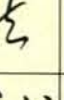
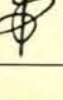
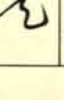
Non ci sono motivi per affermare che il cinese sia più o meno ambiguo di altre lingue. In una frase in lingua inglese come la seguente *It is too hot to eat* (è troppo caldo per mangiare), il pronome *it* può riferirsi o al tempo o al cibo o all'animale che si accinge a mangiare. *Hot* può inoltre significare « contenente molto calore » (in cinese *tàng*) oppure « piccante » (in cinese *là*).

La lingua cinese come oggetto di studio risale all'indietro fino all'inizio dell'era cristiana. A quell'epoca i cinesi avevano già prodotto opere raffinate di dialettologia e classificazione semantica. Le basi per lo studio storico dei suoni cinesi furono gettate nel seicento e nel settecento dai grandi eruditi della dinastia Ch'ing (grafia ufficiale moderna: Qing) su cui i linguisti cinesi moderni si basano.

La chiave per ricostruire i suoni di una lingua dei secoli passati risiede nella pronuncia attuale. Il metodo fondamentale consiste nel confrontare la pronuncia di morfemi in dialetti contemporanei e inferirne quale possa essere stata la pronuncia originaria. L'inferenza non è una congettura ingiustificata, ma si fonda sui documenti disponibili e sulla conoscenza dei princípi linguistici generali che stanno alla base del mutamento dei suoni.

Data la natura non fonetica del sistema di scrittura cinese, la ricostruzione di come una lingua era parlata molti secoli fa può sembrare un compito impossibile. Qualcuno potrebbe pensare che sia molto più facile ricostruire come venivano pronunciati i suoni di lingue scritte con l'aiuto di un alfabeto, come la lingua slava ecclesiastica antica o il sanscrito; in realtà il compito presenta difficoltà anche in questo caso, non esistendo un modo diretto di determinare in che modo una lettera venisse pronunciata. Anche nel caso di lingue alfabetiche si perviene dunque ai valori fonetici solo per inferenza.

Il cinese ha, rispetto ad altre lingue, un grande vantaggio costituito da una abbondanza di documenti scritti antichi che risalgono a un'epoca anteriore a quella della letteratura di qualsiasi altra lingua nel mondo. Il fatto che la forma dei caratteri cinesi non sia molto influenzata da mutamenti nella pronuncia è piuttosto un vantaggio, in quanto aiuta i linguisti a stabilire quali morfemi siano legati etimologicamente.

Lo studio dei dialetti cinesi è stato ostacolato in passato da una sovrabbondanza di dati, che rendeva scomode e interminabili le ricerche. L'avvento di grandi cervelli elettronici ha facilitato la manipolazione dei dati. Chin-Chuan Cheng e io, con l'aiuto di vari colleghi, abbiamo sviluppato un dizionario dei dialetti su un computer, da noi chiamato DOC. Il programma viene attualmente elaborato ai centri elettronici dell'Università di California a Berkeley e dell'Università di Illinois. Esso include la pronuncia di più di 2000 morfemi in ciascuno di 20 dialetti cinesi.

Il nostro lavoro con DOC ci ha consentito finora di accumulare prove del fatto che i mutamenti linguistici procedono in modi essenzialmente paralleli a quelli vigenti nell'evoluzione biologica, qual è stata definita da Charles Darwin nell'*Origine dell'uomo*. In entrambi i casi il meccanismo del mutamento risiede nella variazione. Quando appaiono due o più varianti, le forme rivali entrano in competizione per la sopravvivenza. Per esempio, nella pronuncia americana, la vocale di *room* varia tra quella di *pool* e quella di *put*, mentre la *s* di *disobey* varia tra una *s* e una *z*. Più valide nel meccanismo della selezione, che è costante nel tempo come nelle lingue, sono quelle forme che possono essere pronunciate e percepite con maggiore facilità. Questo vantaggio selettivo determina la sopravvivenza di una forma a danno dell'altra.

Con la programmazione di computer come DOC siamo stati in grado di vagliare grandi quantità di dati per individuare mutamenti di suoni di interesse particolare. Alcuni mutamenti stanno iniziandosi solo ora, altri sono a metà del loro corso, altri ancora si trovano nella fase terminale. Nel dialetto cháozhou, per esempio, c'è un mutamento da un tono all'altro che ha interessato finora circa la metà del vocabolario. Avendo accesso a un gran numero di tipi ben definiti di mutamento di suoni, siamo ora in grado di studiare con molta maggior efficacia il vantaggio selettivo che determina la direzione dell'evoluzione linguistica.

Altri aspetti della lingua cinese sottoposti a un intenso studio con l'aiuto di computer sono la morfologia e la sintassi. Con la collaborazione di Stephen W. Chan, Benjamin K. Tsou e altri abbiamo sviluppato un dizionario meccanico di più di 70 000 voci, insieme con i necessari programmi per la traduzione dal cinese all'inglese. Una buona indicazione della misura in cui comprendiamo la struttura di una lingua è data dal modo in cui possiamo scomporla sì da rendere possibile una traduzione meccanica. Pur essendo ben lontani dall'essere in grado di tradurre in inglese una poesia T'ang senza perderne la squisita sensibilità, possiamo invece tradurre con una precisione ragionevole testi scientifici.

Ora che le relazioni tra la Repubblica Popolare cinese e le nazioni occidentali stanno avviandosi a una maggiore normalità, l'interesse per la lingua cinese sta crescendo a un ritmo sempre maggiore. Se consideriamo che questa lingua è quella parlata nel mondo dal maggior numero di persone e che la sua letteratura è quella più antica, questo interesse è molto tardivo. Col crescere di tale interesse possiamo considerare la possibilità di sondaggi più approfonditi nella storia e nella struttura della lingua e nell'influenza che la lingua ha avuto sullo sviluppo culturale e intellettuale del popolo cinese. Questi studi consentiranno sicuramente anche una migliore comprensione generale della natura del linguaggio umano.

Sistemi telesensori spaziali e risorse terrestri

L'attuale tecnologia delle teleosservazioni dai satelliti fornisce utili informazioni per lo studio e la gestione delle risorse terrestri

di Luigi G. Napolitano

La tecnologia, nei suoi molteplici sviluppi e nelle sue implicazioni consumistiche, è stata spesso posta sul banco degli accusati in quanto ritenuta diretta responsabile delle azioni più nocive all'equilibrio ecologico. Quasi paradossalmente, la stessa tecnologia viene altrettanto spesso biasimata in quanto ritenuta incapace di affrontare e risolvere i gravissimi problemi derivanti dalla immanente scarsità di molte risorse terrestri – inadeguate a sopperire ai sempre crescenti bisogni di una umanità in espansione – e dalle inerenti difficoltà di gestirle in modo da perequare situazioni di sottosviluppo per scarsa accessibilità ai beni elementari di consumo.

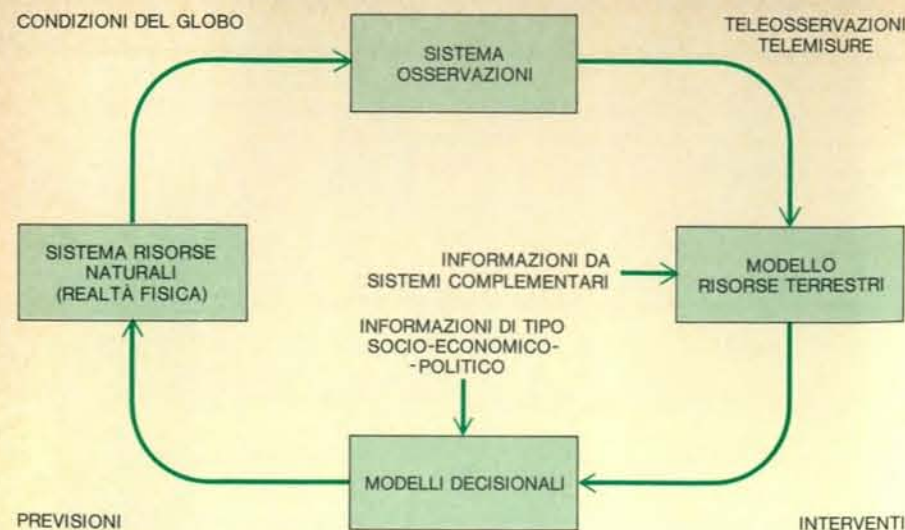
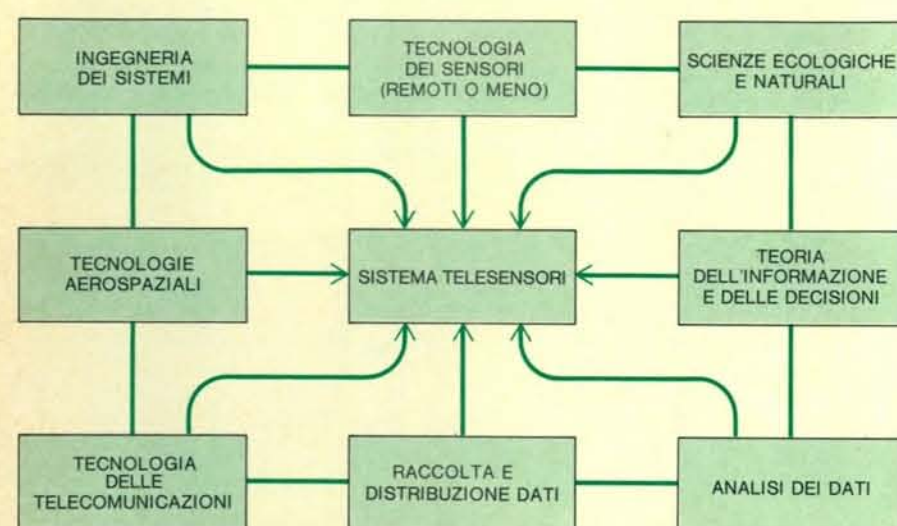
La disputa che, a Stoccolma, durante la Conferenza dell'ONU sull'ambiente, ha a volte assunto toni drammatici, è tutt'ora in corso. C'è chi sostiene, d'accordo con Barry Commoner, che il deterioramento ecologico è essenzialmente dovuto all'inquinamento di cui è responsabile primo « il cattivo » uso della tecnologia. Altri, come Paul R. Ehrlich, ritengono che le cause della crisi ecologica siano molto più complesse, antedatino in molti casi lo sviluppo tecnologico e vadano ricercate nella complessa interazione di tre fattori principali: crescita della popolazione, aumento non perequato né uniforme del benessere, sviluppo tecnologico. Tutti sono però d'accordo con Commoner quando afferma che: « La civiltà comporta una serie di processi ciclicamente interdipendenti. La più parte di essi ha una intrinseca tendenza alla crescita, ma non le risorse naturali, insostituibili e assolutamente essenziali, della Terra e dell'ecosfera ».

Il problema della gestione di queste risorse va affrontato su scala globale utilizzando tutte le tecniche e metodologie più avanzate che la scienza e la tecnologia mettono a disposizione. Da innumerevoli dibattiti, da diverse e diversificate disamine, da varie dispute, è infatti emerso, inconfutabile, il carattere di globalità rivestito dai problemi inerenti all'ecosfera e alla biosfera e da quelli connessi con la loro interazione. Alla globalità delle problematiche deve corrispondere la globalità dell'impostazione e, conseguentemente, la globalità delle « osservazioni ». Le nuove dimensioni dei problemi posti dal deterioramento ecologico, dalla dissipazione delle risorse naturali e dalla espansione della popolazione, richiedono pertanto un nuovo modo di « vedere » la Terra, più adeguato ed efficace di quello possibile con i mezzi convenzionali.

Un sistema globale per la gestione delle risorse terrestri deve contenere per lo meno gli elementi indicati nella figura in alto a pagina 82. I modelli matematici che devono fornire informazioni e previsioni circa la fenomenologia della ecosfera e della biosfe-

SETTORI E OBIETTIVI PRINCIPALI DEL RILEVAMENTO A DISTANZA

SETTORE	OBIETTIVI
AGRICOLTURA E FORESTE	Ricognizione (uso terreni, suolo) Individuazione colture Rivelazione incendi Inventari forestali
GEOGRAFIA	Mappe; scale da 1 : 25 000 a 1 : 1 000 000 Ricognizioni (aree industriali, urbane, trasporti)
GEOLOGIA	Mappe geologiche; scale 1 : 250 000 e 1 : 1 000 000 Mappe termiche Prospezioni (minerali, petroli, metalli pesanti) Studi fenomeni dinamici (eruzioni, terremoti)
IDROLOGIA	Mappe (bacini fluviali; alluvioni) Ricognizione (acqua, neve, ghiaccio, umidità terreni, sedimentazioni, erosioni) Controllo stazioni idrologiche
OCEANOGRAFIA	Mappe (oceaniche, costali) Raccolta dati da boe Sorveglianza ghiacci Localizzazione pesci Ricognizione (stato dei mari, opere ingegneristiche costali, pericoli per la navigazione, inquinanti marini)
SCIENZE ATMOSFERICHE	Osservazioni (nubi, tempeste) Misure (velocità del vento, temperatura, densità, umidità, ozono, bilancio termico) Distribuzioni globali (precipitazioni, inquinamento atmosferico)

Nella tabella sono elencati i principali settori in cui è possibile applicare il rilevamento a distanza, assieme a una sommaria descrizione dei principali risultati ottenibili.

Rappresentazione schematica del flusso di informazioni in un sistema globale per la gestione delle risorse terrestri (*sopra*) e schema a blocchi utilizzato per mostrare le mutue relazioni tra le diverse scienze, le diverse tecnologie e i sistemi telesensori (*sotto*).

ra hanno bisogno di dati ottenuti da osservazioni e misure effettuate con i mezzi piú opportuni e adeguati. Solo cosí potranno risultare attendibili e valide le strategie manageriali dedotte da modelli decisionali che amalgamano e armonizzano dette previsioni con quelle desunte dall'analisi dei sottosistemi socioeconomici, politici, culturali. Il ciclo si chiuderà con le susseguenti osservazioni delle nuove condizioni risultanti dall'effetto combinato della realizzazione delle strategie scelte e dei fenomeni naturali.

Il rilevamento a distanza

La scienza e le tecnologie spaziali, originariamente sviluppate per esplorare lo spazio e raggiungere la Luna e i pianeti a noi piú vicini, possono apportare un contributo notevole alla efficace realizzazione di un tale schema gestionale. I satelliti artificiali (e, in un futuro abbastanza prossimo, le piattaforme spaziali) possono raccogliere direttamente, a mezzo di telesensori, o indirettamente, rilevandoli da reti di sensori posti sulla superficie terrestre, dati globali sulle condizioni dell'atmosfera, degli oceani, della superficie terrestre, sulle attività dell'uomo e degli altri organismi viventi e sulle loro mutue interazioni e influenze. I telesensori già disponibili mettono l'uomo in grado di osservare « dall'alto » e con una sensibilità percettiva non piú limitata al solo spettro visivo: la globalità dell'osservazione è cosí ulteriormente potenziata dalla maggiore « capacità » di osservazione.

I sistemi di teleosservazioni e telemisure spaziali (che ovviamente non rimpiazzano quelli piú o meno convenzionali, bensí li complementano, integrano e perfezionano) rappresentano una vera e propria « rivoluzione » sia per la natura che per il volume dei dati raccolti. Va subito detto però che queste caratteristiche inducono notevoli problematiche, alcune delle quali spesso non sufficientemente riconosciute o non ancora soddisfacentemente risolte, relative alle altre fasi del sistema gestionale dell'informazione: trasmissione, trattamento, elaborazione e preanalisi dei dati, immagazzinamento e disponibilità, distribuzione, analisi dei dati, (definizione e ricerca dei contenuti informativi), estrazione delle informazioni, loro utilizzazione.

I sistemi telesensori spaziali sono ancora in una fase di sviluppo pre-operativo avente per obiettivo la conferma della loro fattibilità tecnologica limitatamente al ruolo « osservazione ». In un prossimo futuro, se, contemporaneamente, sarà anche stata definitivamente provata la loro utilizzabilità pratica e concreta (e, per fortuna, dopo gli USA e l'URSS anche l'Europa sembra voglia seriamente impegnarsi), si potrà pensare agli sviluppi ovvi e naturali. Andrà considerata allora l'integrazione dei sistemi spaziali in un sistema globale di telemisure – che comprenda anche piattaforme telesensorie situate su aerei, palloni, razzi sonda, stazioni a terra – le cui diverse interfacce, non esclusa quella uomo-macchina, vanno coscienziosamente analizzate. Successivamente il ruolo dei satelliti potrà essere esteso da quello di semplice osservazione a quello di osservazione abbinata a funzioni di comando e controllo.

Satelliti artificiali

I satelliti artificiali per applicazioni terrestri possono essere distinti in due grandi classi: quelli che già esplicano « servizi » (satelliti meteorologici, per telecomunicazioni, per controllo del traffico e della navigazione) e quelli, quali i satelliti per lo studio delle risorse terrestri, che raccolgono dati e informazioni per successive elaborazioni sia di tipo scientifico (studio di fenomeni fisici o cicli naturali della biosfera) che di tipo utilitaristico.

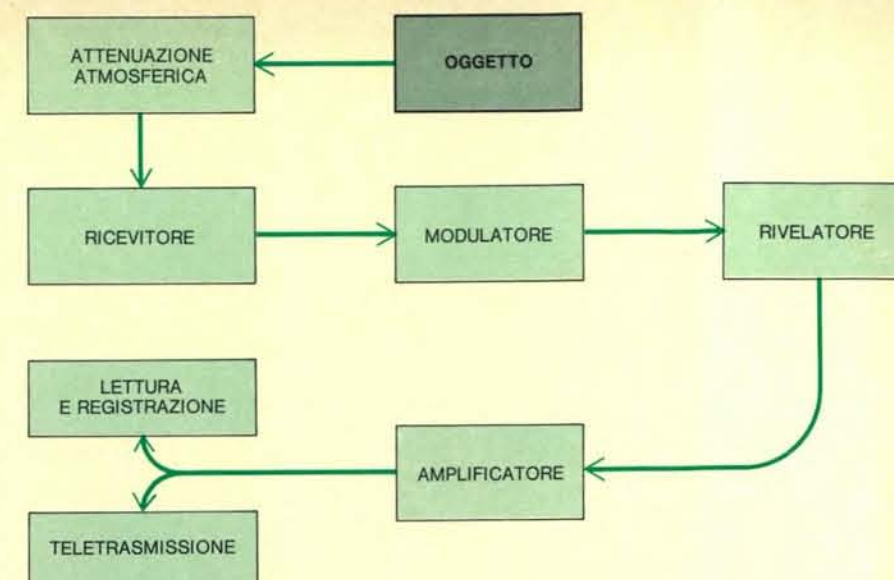
Le telemisure da satelliti e/o piattaforme spaziali, come già accennato, rappresentano solo una delle possibili alternative. Come per qualsiasi altra tecnica, ci saranno delle cose che si potranno fare; ci saranno altri campi in cui essa potrà far le cose meglio (in senso assoluto o relativo, tecnico-scientifico o economico) di come potrebbero essere fatte altrimenti; ci saranno dei campi in cui essa potrà integrare e perfezionare quanto fatto con altre tecniche sia procurando dati e informazioni carenti, sia verificando dati e informazioni ottenuti diversamente; ci saranno infine delle cose che non possono essere assolutamente fatte con i satelliti.

Un discorso largamente qualitativo può però già essere fatto, centrato su due quesiti fondamentali. Quali sono le caratteristiche peculiari delle osservazioni da satelliti e quali i vantaggi specifici che portano a preferirli ad altri sistemi? Quali sono i campi di potenziale applicazione dei satelliti e quali elementi si hanno per poter essere ragionevolmente confidenti sulla possibilità di realizzare queste potenzialità?

Tra le piú importanti caratteristiche peculiari dell'impiego dei satelliti le piú interessanti sono senza dubbio le seguenti: 1) La possibilità di eseguire osservazioni sinottiche di grandi estensioni superficiali il che, per esempio, consente la pratica eliminazione degli errori connessi con la costruzione di « mosaici » e delle difficoltà connesse con la calibrazione di piú strumenti distribuiti su vaste zone. 2) La globalità e la simultaneità delle osservazioni che permettono di affrontare i problemi con un'impostazione sistemistica. 3) La possibilità di eseguire osservazioni reiterate, il che consente di registrare con continuità i fenomeni che variano nel tempo. 4) La possibilità di acquisire informazioni quasi in tempo reale.

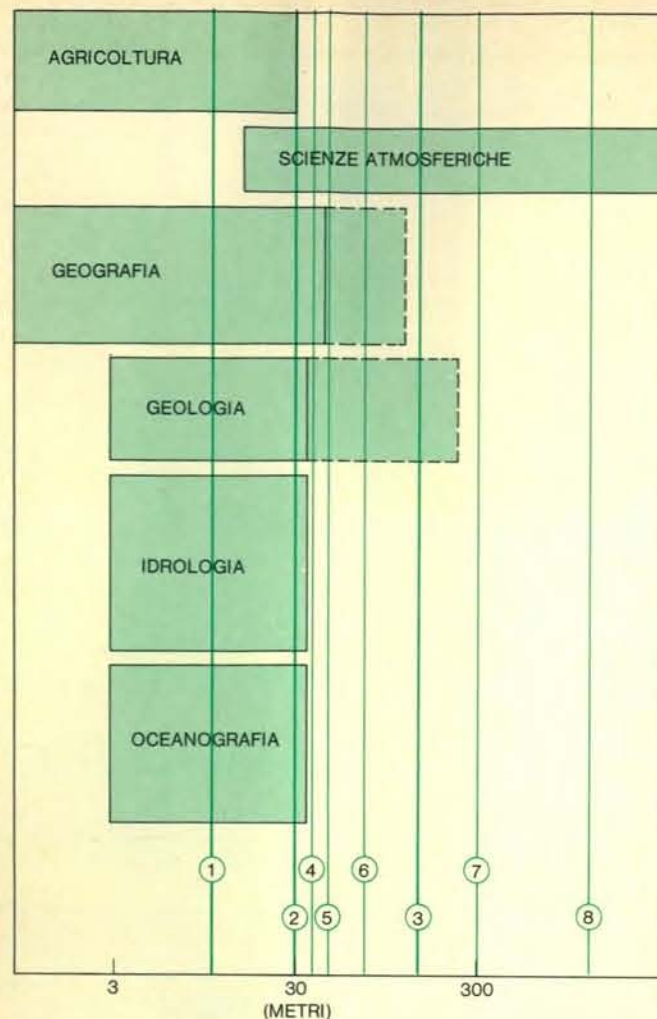
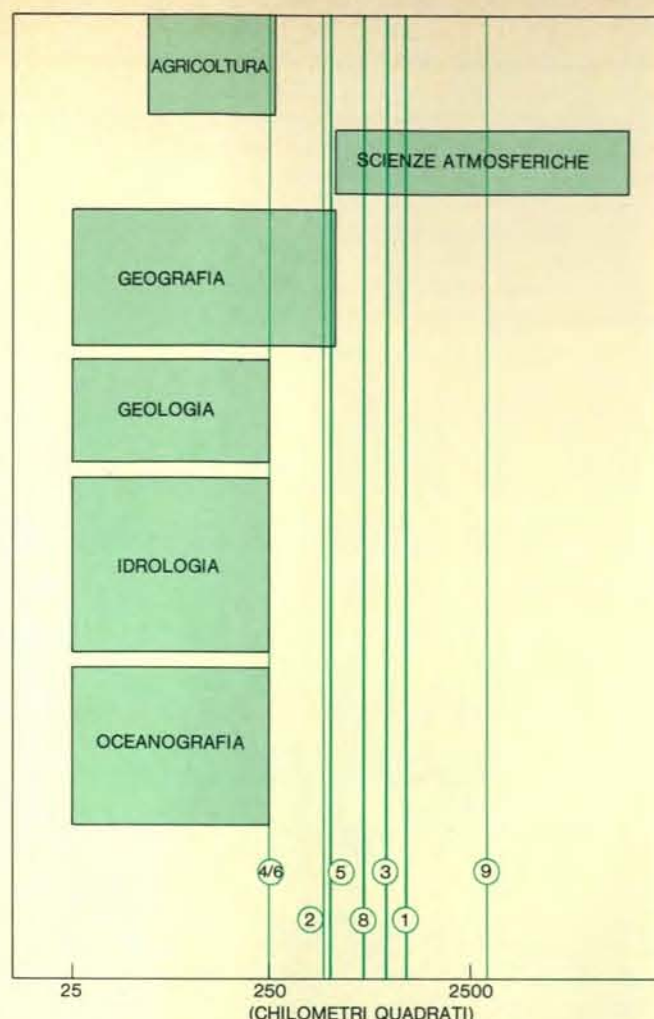
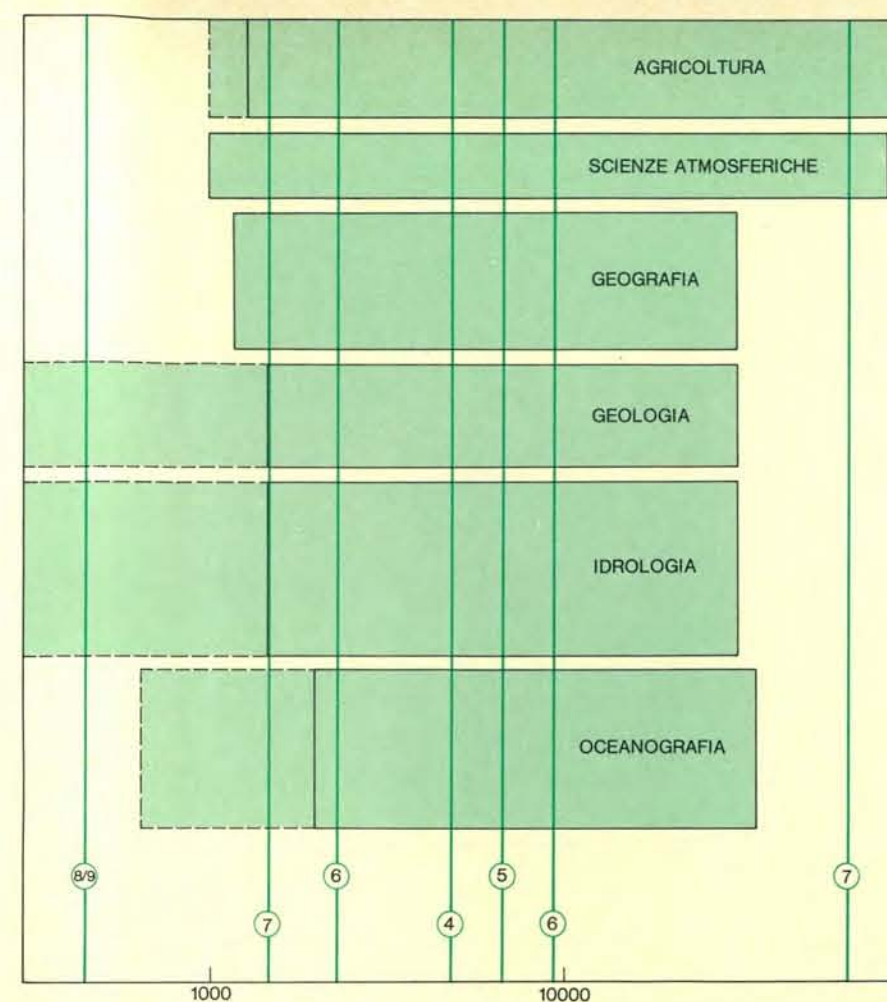
Il quadro va completato mettendo in evidenza l'accuratezza delle osservazioni dovuta alla precisa regolarità del moto del satellite e all'assenza di vibrazioni e la capacità di osservare zone inaccessibili con la conseguente possibilità di ottimizzare il sistema raccolta dati, discriminando eventuali regioni ristrette sulle quali potranno essere in seguito inviati aerei per osservazioni con risoluzione maggiore.

I telesensori

Secondo la definizione nata dal Comitato dell'ONU per gli usi pacifici dello spazio durante la prima riunione del Comitato tecnico (maggio 1971), « un sistema di rilevamento a distanza (*remote sensing system*) è un sistema per misurare le condizioni ambientali mediante sensori elettromagnetici installati su aerei o su satelliti artificiali. Il rilevamento a distanza è una logica estensione delle attuali possibilità fotografiche e un tentativo di misurare mediante una porzione piú ampia dello spettro elettromagnetico ». In questo contesto ci sembra però piú adeguato parlare del rilevamento a distanza come di una scienza che si interessa della raccolta di dati relativi alla superficie terrestre e/o all'ambiente terrestre per mezzo di sensori di vario tipo montati su aerei, satelliti, razzi sonda, palloni, ecc. e della loro analisi per ottenere informazioni utili per capire e gestire l'ambiente in cui vive l'uomo.

Questo schema a blocchi rappresenta, nelle sue linee essenziali, il funzionamento dei sensori passivi a immagini, cioè macchine fotografiche e televisive e sistemi Vidicon.

SISTEMI OTTICI A IMMAGINI

	SENSORE	REGIONE SPETTRALE	CAMPO VISIVO (RADIANTI)	RISOLUZIONE (MILLIRADIANTI)
MACCHINE FOTOGRAFICHE	f = 15 cm	0,4 — 1,0 cm	1,29	0,23
	f = 30 cm	0,5 — 0,7 cm	0,72	0,078
	f = 60 cm Panoramica	0,5 — 0,7 cm	1,43	0,023
	Lunar Orbiter f = 60 cm Panoramica	0,4 — 0,7 cm	0,35	1,64
	Lunar Orbiter f = 7,5 cm		0,76	0,090
	APT		1.89	0,045
SISTEMI TELEVISIVI	AVCTS		0,64	0,052
	1 x 1 Vidicon f = 12,5 cm		0,2	0,38
	2 x 2 Vidicon f = 25 cm			6,8
RADIOMETRI OTTICI	Infrarosso a media risoluzione (MRIR)	0,2 — 3,0 μ	0,45	83
	Infrarosso ad alta risoluzione (HRIR)	3,4 — 4,2 μ	2	8,7
	A tre canali (3 C)	0,64 μ 0,8 μ 10,5 μ	0,14	2,0
	Filtro selettivo	14-16 μ	0,175	175

Nella tabella sono elencati i principali sistemi ottici a immagini attivi e passivi. I sistemi ottici (macchine fotografiche e macchine televisive) sono sistemi attivi, i radiometri sono invece sistemi passivi. Non sono elencati i diversi tipi di radiometri a microonde.

1. M.F. 60 CENTIMETRI (PANORAMICA)
2. M.F. 30 CENTIMETRI
3. M.F. 15 CENTIMETRI
4. 2x2 VIDICON
5. LUNAR ORBITER (PANORAMICA)
6. 1x1 VIDICON
7. LUNAR ORBITER
8. AVCS
9. APT

In queste tre figure sono rappresentati tre diversi tipi di confronto tra i requisiti e le prestazioni dei sensori ottici a immagini. I requisiti sono rappresentati, per ogni disciplina, dalla base del rettangolo a essa corrispondente (le parti tratteggiate corrispondono a requisiti meno stringenti). Le prestazioni sono invece individuate dalle righe verticali (sottili per orbita a 900 km e spesse per orbita a 450 km) e dai numeri che corrispondono alla legenda a sinistra. Il primo diagramma si riferisce alla risoluzione al suolo, il secondo all'area coperta da un'immagine e l'ultimo al rapporto tra dimensione caratteristica dell'area ripresa e risoluzione al suolo.

I telesensori costituiscono l'elemento cruciale del sistema di rilevamento in quanto devono essere in grado di fornire, da altitudini orbitali, le informazioni specifiche interessanti gli utenti. La stragrande maggioranza di sistemi telesensori utilizza in un modo o nell'altro radiazioni elettromagnetiche; le sonde acustiche rappresentano una delle poche eccezioni. Tutti i materiali assorbono, trasmettono, riflettono ed emettono selettivamente radiazioni elettromagnetiche: l'insieme delle proprietà che caratterizzano l'interazione tra l'oggetto e le radiazioni in una data banda dello spettro elettromagnetico è genericamente indicato come « firma » (*signature*) dell'oggetto stesso relativa alla data banda. Queste caratteristiche spettrali contengono informazioni relative a posizione, dimensioni, forma e proprietà fisico-chimiche degli oggetti per cui, analizzando le « firme » relative a uno o piú bande, si possono, in linea di principio, individuare, caratterizzare e discriminare gli oggetti.

I sistemi telesensori possono essere classificati in diversi modi. Due classificazioni di tipo generale li suddividono in sistemi attivi e sistemi passivi, o in sistemi a immagini fotografiche (su film) e sistemi a immagini ottenute con composizione di segnali elettrici e in sistemi non a immagini. Altre classificazioni possono essere fatte in base alle bande dello spettro elettromagnetico in cui i sensori operano o in base ai loro principi di funzionamento.

I sistemi passivi captano e registrano l'energia disponibile naturalmente, emessa o riflessa dal mezzo o oggetto che si vuol osservare. L'esempio classico di un tale sistema è la fotografia con luce naturale. I sistemi attivi invece impiegano una loro sorgente artificiale per « illuminare » l'oggetto da osservare. Un esempio elementare di sistema attivo è la fotografia con il flash, ma quello piú importante è il radar. I sistemi attivi offrono il vantaggio di poter influenzare il segnale di ritorno e di essere, pertanto, meno sensibili a limitazioni imposte dalla energia naturale disponibile. Paragonati ai sistemi passivi essi sono però piú complessi e richiedono una potenza maggiore.

Per lo studio delle risorse terrestri il campo di azione dei sensori è, allo stato attuale, essenzialmente ristretto allo spettro visibile, all'infrarosso e alle microonde. La banda del visibile, con le sue tre classiche suddivisioni blu (0,4-0,5 μ), verde (0,5-0,6 μ) e rossa (0,6-0,7 μ), è la regione delle spettro nella quale operano sia l'occhio umano sia le macchine fotografiche comuni. La banda dell'infrarosso va da 0,7 μ a 1 mm ed è, di solito, suddivisa in quattro bande, anche se questa suddivisione non è accettata da tutti: l'infrarosso vicino (0,7-3 μ, NIR), l'infrarosso medio (3-6 μ, MIR), l'infrarosso lontano (6-15 μ, FIR) e l'infrarosso estremo (15-1000 μ, XIR). La banda delle microonde va dalle onde millimetriche alle metriche.

L'impiego di altre bande dello spettro elettromagnetico, oltre quella del visibile, offre un certo numero di vantaggi. L'infrarosso e le microonde forniscono nuovi tipi di informazioni e le corrispondenti osservazioni sono notevolmente meno condizionate dalla presenza delle nuvole e non richiedono la presenza del Sole. Inoltre, siccome un dato oggetto ha « firme » spettrali diverse a seconda della lunghezza d'onda, certe particolari caratteristiche di un oggetto possono essere discernibili a una lunghezza d'onda ma non a un'altra, per cui dal confronto di immagini prese simultaneamente a diverse lunghezze d'onda può discendere la possibilità di discriminare caratteristiche o fenomeni che non mostrano differenza alcuna nel campo del visibile.

Per esempio, sensori all'infrarosso possono produrre delle « mappe termiche » e dalla loro analisi si possono dedurre le distribuzioni dei gradienti di temperatura. Entrambi questi elementi, specialmente quando si associ a essi la dimensione temporale che permette sia di seguirne l'evoluzione superficiale sia di dedurre informazioni sulle condizioni in profondità, sono di grande utilità nello studio di fenomeni vulcanici, dell'inquinamento delle acque e delle correnti. L'efficace impiego delle teleosservazioni presuppone che siano note le caratteristiche spettrali dell'atmosfera. Qualunque sia la tecnica usata, attiva o passiva, si tratterà sempre, in definitiva, di interpretare un segnale elettromagnetico (la « firma ») che risulta dall'azione combinata delle condizioni dell'oggetto « osservato » e delle modifiche subite dalla somma delle radiazioni emesse e riflesse lungo il suo percorso attraverso l'atmosfera.

L'atmosfera della Terra, di per sé, assorbe, emette, diffrange e diffonde energia elettromagnetica in modo diverso alle diverse lunghezze d'onda. Ne consegue che avranno significato operativo solo quei sensori agenti in quelle particolari bande in cui l'atmosfera permette la trasmissione di energia, almeno in determinate condizioni; in caso contrario sarà necessario distinguere tra le firme in-situ e quelle registrate dai sensori.

Le modifiche subite dalle radiazioni di una certa lunghezza d'onda nel passare attraverso l'atmosfera dipendono dalla composizione del percorso e, in misura minore, dalla distribuzione di pressione e temperatura lungo di esso. Le regioni dello spettro in cui l'assorbimento dell'atmosfera (pura) è ridotto al minimo sono comunemente chiamate « finestre ». Esse chiaramente influenzano la scelta del sistema sensorio passivo e quella delle bande da adoperare nei sistemi attivi.

Nella regione visibile l'atmosfera (pura) trasmette approssimativamente il 90 % della radiazione solare incidente; circa il cinquanta per cento dell'energia solare totale ricevuta sulla superficie terrestre si trova in questa banda. Delle bande di assorbimento, piuttosto deboli, si trovano verso l'estremo superiore delle lunghezze d'onda e sono dovute principalmente all'ozono, all'ossigeno molecolare e al vapor d'acqua. Una certa importanza rivestono anche gli effetti di diffusione molecolare e di aerosol.

Nelle prime tre bande dell'infrarosso i costituenti piú attivi della atmosfera sono il vapor d'acqua (la cui presenza è limitata alla troposfera – approssimativamente i primi 14 chilometri – ma la cui concentrazione varia molto con la posizione e il tempo), l'anidride carbonica (distribuita abbastanza uniformemente – approssimativamente 320 parti per milione – sino a un'altezza di 100 chilometri; è stato calcolato che il rapporto di mescolamento aumenta di 0,7 ppm per anno), ossido d'azoto e ozono (presenti nella stratosfera e oltre). L'effetto complessivo dell'assorbimento dovuto a questi elementi è che, in queste tre bande, le « finestre » disponibili sono poche e di estensione limitata. Le piú importanti sono quella nel medio infrarosso (4,6-4,8 μ) e quella nel lontano infrarosso (8-12 μ). Va però notato che le curve spettrali di trasmissione attraverso l'atmosfera hanno una struttura complessa e variano con il posto, la stagione, l'ora del giorno, l'angolo visuale, ecc. Al disopra dei 100 μ (infrarosso estremo e microonde) l'atmosfera è abbastanza trasparente: ci sono solo quattro linee di assorbimento piuttosto intense dovute al vapor d'acqua (0,926; 0,928; 1,64; 13,48 mm) – piú alcune altre deboli – e una linea d'assorbimento a 2,5 millimetri, piú una banda

1. MRIR
2. 3C
3. HRIR

In queste due figure sono rappresentati due diversi tipi di confronto tra i requisiti e le prestazioni dei radiometri ottici. Anche in questo caso i requisiti per ogni disciplina sono rappresentati dalle basi dei rettangoli e le prestazioni dalle righe verticali. Il primo diagramma si riferisce alla risoluzione al suolo, il secondo alla copertura.

d'assorbimento a 5 millimetri dovute all'ossigeno molecolare.

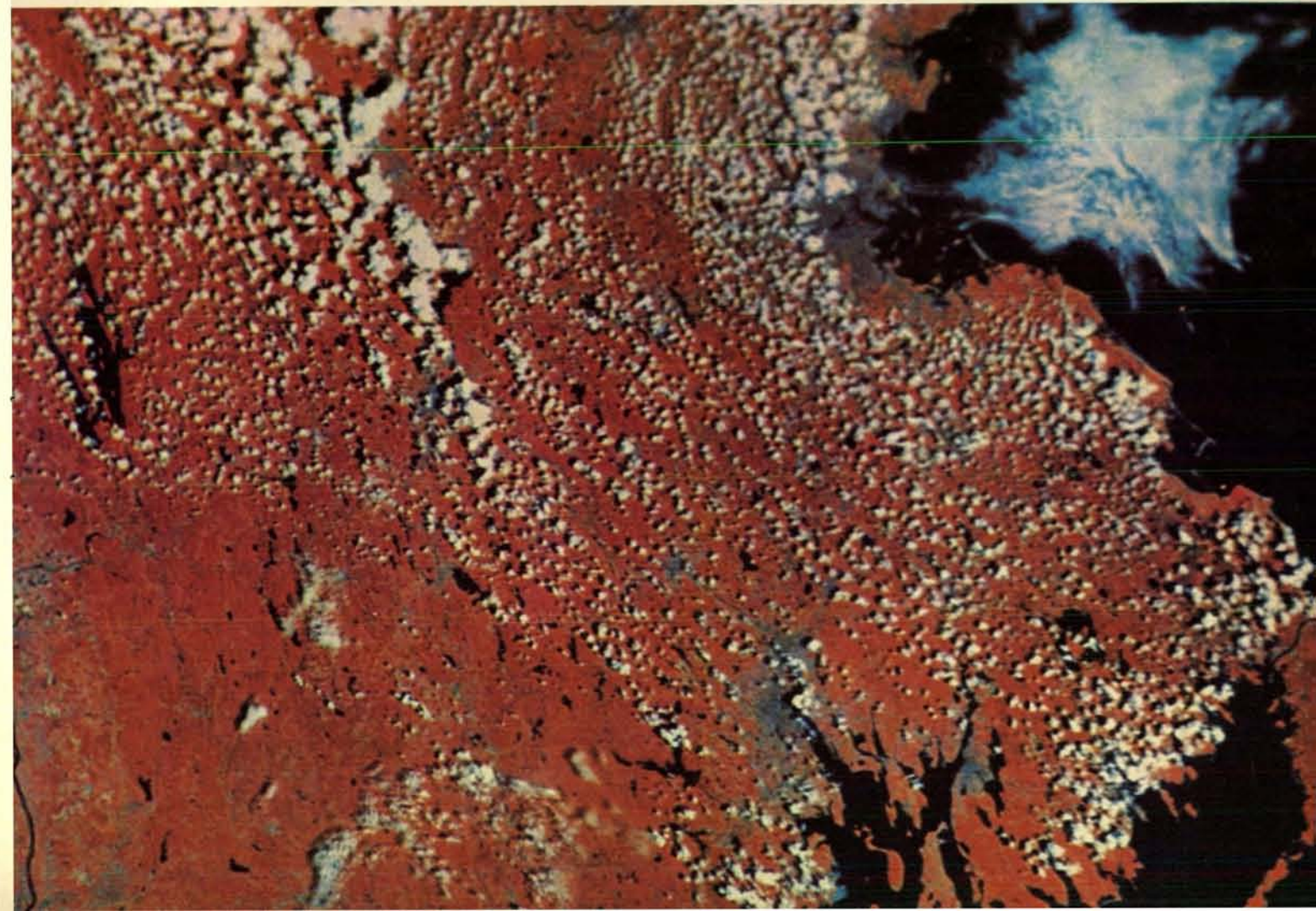
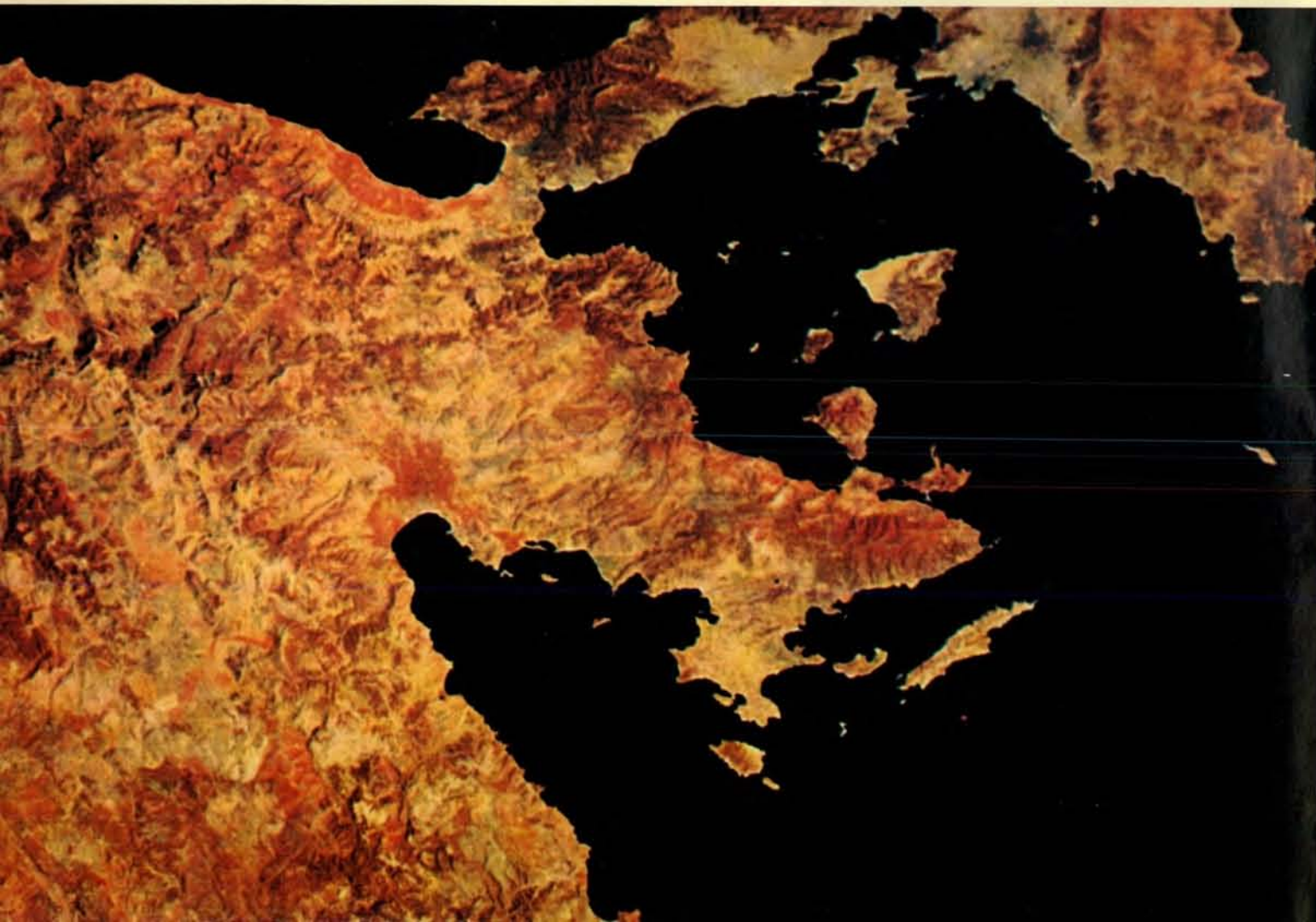
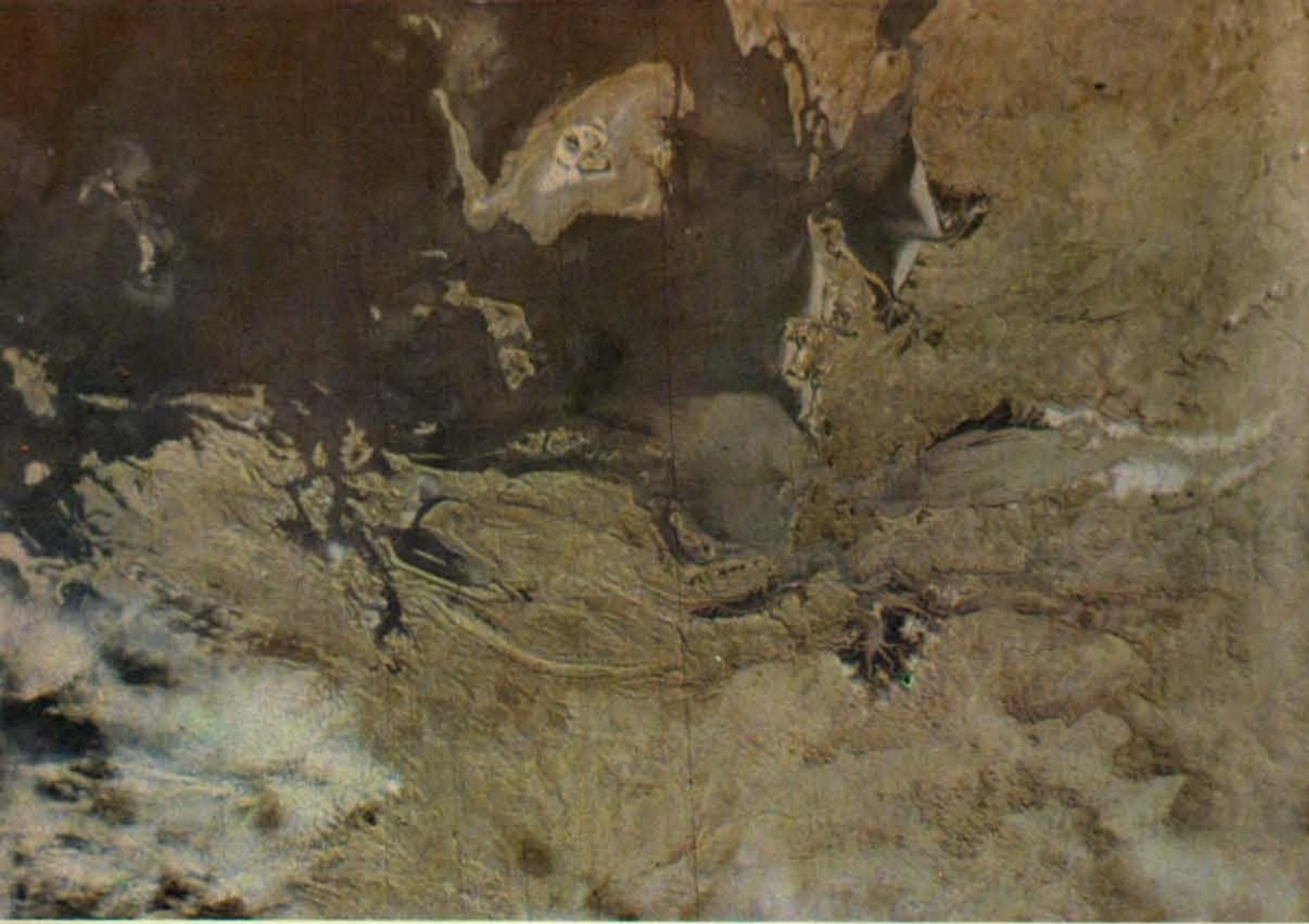
Infine, poiché la maggior parte delle nubi si trova nella troposfera, non impone forti limitazioni alle misure delle proprietà della stratosfera. Il discorso è ben diverso, invece, per osservazioni della superficie terrestre perché l'energia a lunghezze d'onda infrarosse e minori viene sia diffusa sia assorbita dalle gocce d'acqua; in questo caso sono necessarie sia una risoluzione spaziale sufficientemente elevata per poter vedere attraverso le aree libere tra le nubi, sia osservazioni indipendenti per individuare la presenza di nubi in una parte del campo visivo.

Sistemi passivi

I sistemi a immagini ottiche sono, nella loro forma più semplice, le macchine fotografiche metriche o multispettrali a film. Gli elementi che influenzano la qualità delle immagini fornite dai sistemi fotografici a film sono fattori fisici, quali l'illuminazione e la riflettanza degli oggetti, la diffrazione, l'assorbimento e la diffusione atmosferici, l'angolazione dei raggi solari (che influenza quantità e qualità spettrale dell'energia captata dal sensore); le caratteristiche dei film (sensitività, granularità, ecc.); i tipi di film (bianco e nero, all'infrarosso, a colori) e le loro caratteristiche fisiche (come reagiscono a sollecitazioni e come sono influenzati da basse pressioni, radiazioni, umidità, temperatura); tipo e proprietà dei filtri; condizioni cinematiche e dinamiche della piattaforma.

Tra i vantaggi dei sistemi fotografici si possono annoverare la risoluzione superiore a quelle di altri sistemi, anche televisivi o a scansione ottico-meccanica; i limitati interventi di rettificazione (specie con le macchine fotografiche di alta precisione) e la relativa facilità di analisi delle fotografie; i costi e le potenze impegnate relativamente bassi. A questi vantaggi fanno riscontro gli svantaggi dovuti all'impossibilità di impiego notturno o in condizioni meteorologiche avverse; alla maggiore complessità del sistema di teletrasmissione e alle difficoltà di immagazzinamento; alla laboriosità e complessità di impiego di procedimenti automatici di discriminazione e analisi; al campo spettrale di sensibilità limitato al vicino ultravioletto, al visibile e all'infrarosso vicino; alla natura latente dell'immagine e alla non linearità della risposta con conseguente difficoltà di registrare fedelmente sorgenti deboli. Gran parte di questi inconvenienti sono superati mediante l'impiego di camere televisive (anche multispettrali) e sistemi Vidicon (già largamente usati per trasmettere « fotografie » della Luna, di Venere e di Marte) che usano opportune tecniche di scansione elettronica per trasmettere le immagini riprese.

L'insieme delle operazioni di ricezione dei sistemi passivi a immagini non fotografiche può essere rappresentata, nelle sue grandi linee, dallo schema a blocchi illustrato nella figura a pagina 83. La radiazione proveniente dall'oggetto, attenuata dall'atmosfera non perfettamente trasparente, va al ricevitore che la invia al rivelatore che, a sua volta, la converte in un segnale elettrico. Prima di raggiungere il rilevatore la radiazione può passare attraverso un modulatore ottico che serve a

Nella pagina a fronte in alto è mostrato il territorio nord-occidentale del Canada ripreso, da una quota di 915 chilometri, dallo scanner multispettrale montato a bordo dell'ERTS 1. La fotografia è la combinazione di tre fotogrammi (nel verde, nel rosso e nell'infrarosso) e mette in evidenza particolare le caratteristiche della flora. Sono ben visibili i meandri del fiume Mackenzie (*a sinistra*) e ghiacci galleggianti sul mare (*a destra*). La fotografia in basso mostra invece la parte meridionale della Nuova Inghilterra ottenuta con la stessa tecnica. Risultano nettamente visibili le nuvole in direzione di Boston (*in alto a destra*), chiara indicazione di una « zona calda », cioè di un agglomerato urbano.

fornire informazioni per la differenziazione dell'oggetto da dettagli non importanti sullo sfondo. Il segnale elettrico dal rivelatore passa poi all'amplificatore e di qui passa al registratore o viene teletrasmesso.

Le immagini nell'infrarosso possono essere ottenute fotograficamente (nel vicino infrarosso, poiché i film non sono sensibili a lunghezze d'onde superiori a circa 1 μ) o a mezzo di sistemi a scansione ottico-meccanica (piú frequentemente usati di quelli a scansione elettronica per i quali, per lunghezze d'onda superiori a 1,2 μ, le velocità di formazione di immagini tendono a essere piuttosto basse). I rilevatori che trasformano la radiazione infrarossa incidente in segnale elettrico, possono essere di tipo termico (lenti, ma con ampie bande di sensibilità) o di tipo a fotoconduzione (molto piú veloci ma con limitata risposta e richiedenti, nella maggior parte dei casi, raffreddamento). La tecnica degli scanner infrarossi ha raggiunto livelli tali che la qualità delle loro immagini è abbastanza prossima a quella dei sistemi fotografici. Esistono anche sistemi a scansione operanti nello spettro visibile, essenzialmente dello stesso tipo degli scanner a infrarosso. Essi vanno necessariamente impiegati, quando si vogliano sistemi multispettrali includenti bande infrarosse, per garantire che, per ciascuna banda, la scena « osservata » e le sue caratteristiche geometriche siano le stesse in ogni istante e che i tempi di osservazione siano gli stessi.

Allo stato attuale, i sensori passivi a microonde non sono altrettanto ben sviluppati quanto quelli ottici o a infrarosso. Siccome la radiazione emessa nel campo delle microonde è molto minore di quella emessa nella banda infrarossa bisogna far uso di rivelatori estremamente sensibili e ridurre al minimo il rumore interno. La risoluzione è minore di quella dei sensori ottici e a infrarosso. La temperatura registrata da un telesensore a microonde (cosiddetta temperatura di brillanza) dipende dalla temperatura dell'oggetto, dalle radiazioni incidenti su di esso, dalla sua emittanza, riflettanza e trasmittanza (in genere funzioni del coefficiente di assorbimento del materiale, della sua configurazione e forma, dell'angolatura sotto cui è visto e della sua struttura superficiale); dalle vicissitudini, associate all'attraversamento dell'atmosfera terrestre da parte dell'energia emessa, riflessa e trasmessa dall'oggetto; dalle caratteristiche del sistema ricevente.

La spettrometria, infine, è una tecnica ancora relativamente poco usata. Qualche utilizzazione la si comincia a vedere in oceanografia e in geologia, ma soprattutto nel campo dell'inquinamento. La maggior parte degli inquinanti atmosferici ha intensi spettri nell'infrarosso e, poiché questa tecnica è alla base di molti procedimenti di analisi di laboratorio e di controllo di qualità, molti spettri di riferimento sono stati accuratamente misurati e registrati.

Sistemi attivi

Tra i sistemi attivi vanno annoverati, oltre ai già citati radar convenzionali, i radar a apertura sintetica (SAR, *synthetic aperture imaging radar* o anche SLR *side-looking radar*), i radar altimetri e scatterometrici e i laser-altimetri. Le prestazioni dei radar dipendono essenzialmente dai parametri della emittente (lunghezza d'onda, polarizzazione e direzione), dalle proprietà delle superfici irradiate (proprietà dielettriche e conduttive, scabrosità della superficie, risonanze, inclinazione, effetti subsuperficiali, area di diffusione) e da effetti atmosferici (dovuti per lo più agli eco causati da nubi e pioggia, peraltro facilmente discriminabili).

Tra i parametri che possono essere usati per l'identificazione e/o la discriminazione i piú importanti sono la polarizzazione e le caratteristiche temporali e spettrali della radiazione captata dal sensore. Le proprietà riflettenti della maggior parte delle superfici materiali sono sensibilmente affette dalla polarizzazione alle frequenze proprie delle microonde. Variazioni temporali possono dipendere da variazioni (diurne o stagionali) della temperatura propria dell'oggetto, delle condizioni atmosferiche e delle proprietà di riflessione e di emissione dell'oggetto. Nelle immagini fornite dai radar e dai sistemi passivi a microonde la risoluzione è minore che in quelle fornite da sistemi a infrarosso o fotografici a causa della maggiore lunghezza d'onda.

Tra i vantaggi generici dei sistemi radar vanno ricordati la rappresentazione di aree estese con risoluzione praticamente uniforme e la possibilità di controllare la polarizzazione della radiazione emessa. A questi si aggiungono, per radar operanti a lunghezze d'onda dell'ordine di centimetri o metri, in particolare per il SAR, quelli connessi con la possibilità di operare con condizioni atmosferiche qualsiasi, con la possibilità di penetrare il fogliame degli alberi e quindi studiare le caratteristiche del sottobosco e con la possibilità di discriminare oggetti quali nubi o ghiaccio nel mare (che di solito appaiono confusi sulle immagini fotografiche), capacità discriminatorie non realizzabili con altre lunghezze d'onda.

I radar scatterometrici possono infine fornire informazioni sullo stato del mare e sulle grosse asperità del terreno. Quelli convenzionali servono invece per « mappare » fenomeni atmosferici (nubi, precipitazioni). I laser altimetrici possono essere impiegati per ottenere misure dirette di piccole variazioni di altezza (da 30 a 150 centimetri) con piccole risoluzioni a terra. Allo stato attuale ben poco è però stato fatto relativamente all'uso di sistemi attivi da piattaforme orbitanti.

Confronto requisiti-prestazioni

La interdipendenza tra gli obiettivi specifici di una certa missione e le caratteristiche del corrispondente sistema satellite-telesensori è complessa e multiforme. I parametri orbitali, per esempio, influenzano direttamente la capacità dei sensori di fornire osservazioni ai livelli di qualità richiesti e sono collegati alla vita utile in orbita, al tempo necessario per completare la copertura del globo terracqueo, alla frequenza delle osservazioni, alla loro ripetitività e cosí via. Altre interfacce tra veicolo spaziale, infrastrutture a terra e obiettivi della missione sono quelle connesse con la gestione dei dati, il tempo necessario per trasmettere i dati alle stazioni al suolo, le operazioni di lancio, le caratteristiche della propulsione ausiliaria necessaria per mantenere l'orbita richiesta, il controllo di assetto, necessario per mantenere il puntamento dei sensori verso la Terra entro limiti di tolleranza accettabili, e cosí via.

Grosso modo, la progettazione di un sistema satellite-telesensori necessario per compiere determinate missioni nel campo delle risorse terrestri si articola nelle seguenti quattro fasi essenziali.

Nella pagina a fronte sono mostrate due altre fotografie a colori riprese dall'ERTS 1 con la tecnica multispettrale illustrata per le due fotografie precedenti; anche in questo caso si tratta di porzioni dei fotogrammi originari che sono quadrati e un esempio dei quali è dato, in bianco e nero, a pagina 92. Nella fotografia in alto è ben visibile un golfo con un'isola; in particolare evidenza sono poi le caratteristiche geologiche del terreno con catene montuose costiere. Anche nella fotografia in basso, che mostra la Grecia (Atene in alto a destra, l'istmo di Corinto in alto al centro) sono evidenziate le caratteristiche geologiche; il colore giallastro indica aridità e carenza di vegetazione.

Queste due fotografie mostrano la stessa zona (una parte della California settentrionale) ripresa in due diverse bande spettrali dall'ERTS 1. La ripresa in bande diverse consente di evidenziare caratteristiche diverse della zona ripresa. Per esempio nella foto in alto viene evidenziato un fiume (*a sinistra*) praticamente invisibile in quella sotto.

1) Individuazione e caratterizzazione dei requisiti richiesti in funzione degli obiettivi specifici della missione (prestazioni richieste da coloro che devono utilizzare i dati raccolti). 2) Confronto tra detti requisiti e le prestazioni dei sensori e individuazione delle possibili alternative. (In fase programmatica questa analisi servirà anche a individuare le tecnologie che devono essere ulteriormente sviluppate per poter ottenere le prestazioni richieste.) 3) Caratterizzazione dei parametri orbitali e delle tipologie di satelliti e definizione dei vincoli da essi indotti nelle prestazioni dei sensori. 4) Analisi degli elementi comuni (*commonality analysis*) per ridurre al minimo i diversi tipi di sensori che vanno impiegati e i diversi tipi di orbite e di satelliti necessari.

Ci sembra interessante dare un'idea dello stato attuale delle possibilità pratiche dell'impiego di telemisure da satelliti per lo studio delle risorse terrestri, riportando i risultati di uno studio, pubblicato recentemente, sul confronto requisiti-prestazioni.

Lo studio considerava un gruppo di 55 obiettivi, indicativi delle possibili applicazioni nei settori agricoltura e foreste, geografia, geologia, idrologia, oceanografia e scienze atmosferiche. Il numero di tali obiettivi, per ciascun settore, e una loro descrizione sommaria sono riportati nella tabella a pagina 81. Il confronto requisiti-prestazioni è qui limitato ai due sistemi sensori attualmente piú richiesti per operazioni da satelliti: i sistemi ottici a immagini (richiesti per circa il sessanta percento degli obiettivi descritti nella tabella citata) e i sistemi radiometrici passivi a immagini (4 %). Le percentuali per gli altri tipi di sistemi sono nell'ordine: sistemi attivi (10 %); sistemi spettrometrici (8 %), altri sistemi, inclusi quelli basati sulla raccolta di dati da reti sensorie sulla superficie terrestre, (8 %). I sensori considerati e le loro prestazioni sono riportati nella tabella a pagina 83; tutti questi sistemi sono attualmente operativi.

I parametri di confronto per i sistemi ottici a immagini concernono essenzialmente la risoluzione al suolo e la copertura superficiale, caratteristiche, queste, tra loro contrastanti in quanto, ovviamente, una può essere aumentata a spese dell'altra. Va però notato che sono in genere richiesti valori relativamente elevati della copertura superficiale a causa dell'alto costo del processo di composizione a mosaico e delle inevitabili perdite di informazione e riduzione di precisione, presenti anche nei mosaici meglio eseguiti, che tendono ad annullare i vantaggi della osservazione sinottica.

La risoluzione piú frequentemente richiesta si aggira sui 300 metri (*si veda la figura a pagina 84*). Risoluzioni minori, dell'ordine dei 1000 metri, trovano ancora un loro campo di utilizzazione specie in quei casi in cui le informazioni devono essere utilizzate per l'approntamento di mappe meno dettagliate. Dal confronto requisiti-prestazioni si deduce che, per un raggio orbitale di 400 chilometri (tipico di satelliti con uomini a bordo, la cui presenza è necessaria per operare le macchine fotografiche non automatiche), la macchina fotografica con distanza focale di 15 centimetri è adeguata solo per un piccolo numero di obiettivi specifici. Anche se, ovviamente, la situazione migliora con le macchine fotografiche a maggiore distanza focale, non ce n'è nessuna, allo stato attuale, capace di fornire risoluzioni minori di 10 metri (con copertura dell'ordine di 2500 chilometri quadrati). La situazione peggiora per orbite piú alte, quale quella con raggio di 800 chilometri, tipica di satelliti senza uomo a bordo, con vita utile dell'ordine di un anno. Da questa altezza risulta tutt'ora impossibile fotografare aree dell'ordine di 250 chilometri quadrati con risoluzioni migliori di circa 30 metri. Il quadro si presenta nettamente piú favorevole quando il confronto è basato sulle coperture areali, come mostra la figura a destra a pagina 84.

Un altro parametro utile per il confronto requisiti-prestazioni di sistemi ottici a immagini è il rapporto tra una dimensione caratteristica dell'area osservata e la risoluzione al suolo. Questo parametro è importante non solo perché congloba le due caratteristiche contrastanti copertura-risoluzione, ma anche perché permette un confronto in termini di « contenuto di informazione ». Prescrivendo la copertura superficiale (per esempio un quadrato di lato a), la risoluzione al suolo (r) e il numero G di toni o livelli di grigio desiderati, l'utente in pratica prescrive il contenuto di informazione per immagine che è dato (in bit per immagine) da $(a/r)^2 \log_2 G$. In un sistema ottico il rapporto a/r è eguale al rapporto d/q dove d è la dimensione (lato al quadra-

Combinando assieme piú immagini è possibile ricostruire la geografia di intere regioni. Questo esempio si riferisce alla Grecia (un particolare della parte superiore è illustrato a colori a pagina 88); sono ben visibili Capo Matapan (*a sinistra*) e l'isola di Citera. In alto si riconosce l'istmo di Corinto (*al centro*), il Pireo e Atene (*a destra*).

Ecco riportato per intero un fotogramma scattato dall'ERTS 1 il 27 agosto 1972 sulla zona dei golfi di Gaeta (*a sinistra*), di Napoli (*al centro*) e di Salerno (*a destra*). Oltre a formazioni nuvolose e a scie di aviogetti, sono nettamente visibili le isole di Ischia, Procida e Capri. Le diciture ai lati servono a individuare la foto mentre la scala dei grigi in basso viene usata come riferimento.

to) del fotogramma e q la distanza tra coppie di linee. (Una volta fissato un valore tipico del contrasto e assunta una soglia di detettività del fotointerprete, il rapporto d/q dipende dalle caratteristiche del sistema sensorio: macchina fotografica + tipo di film; sistema televisivo).

Il confronto in discorso è riportato in figura a pagina 85. Si nota come, da questo punto di vista, i sistemi a funzionamento automatico, pur essendo discretamente adeguati, sono nettamente inferiori a quelli non automatici (le tre macchine fotografiche). La differenza è sostanzialmente dovuta alle minori dimensioni della superficie sensibile in quanto il potere risolutivo (in coppie di linee al millimetro) dei migliori esemplari delle due classi sono dello stesso ordine di grandezza (per i sistema Vidicon esso è eguale a 88 coppie di linee al millimetro). Il contenuto di informazione di questi sistemi è notevolmente superiore a quello della televisione commerciale a causa sia della migliore risoluzione – da due a quattro volte più grande – sia del maggior numero di toni di grigio – dai 4-8 della televisione commerciale ai 64 dei sistema Vidicon. (Il contenuto di informazione, in bit per immagine, è dell'ordine di $5-8 \times 10^5$ per la televisione commerciale, di $2-8 \times 10^7$ per i sistemi Vidicon e di 10^8-10^9 per i film di più grandi dimensioni usati in aerofotogrammetria).

Un certo numero di obiettivi specifici richiedono la cosiddetta accuratezza metrica, cioè la capacità di individuare la posizione assoluta di oggetti sulla mappa finale con una accuratezza dell'uno per mille. L'accuratezza metrica è indispensabile per le mappe geogra-

Nella figura della pagina a fronte è illustrato schematicamente il flusso di energia e dei vari tipi di informazioni tra il sistema del satellite, le stazioni di controllo, il centro di controllo a terra e, infine, gli utenti.

RAGGI SOLARI
BATTERIE SOLARI
BATTERIE CONVERTITORI DI POTENZA
CONT. TERMICO
COM. ATTITUDINE
TRASMETTITORE RF
TELEMETRIA
COMANDO E CONTROLLO
SISTEMA SATELLITE
INTERFACCIA ELETTRONICA
SENSORI
SUPERFICIE TERRESTRE
STAZIONE RICEVITRICE
STAZIONE DI CONTROLLO
CENTRO ELABORAZIONE DATI
CENTRO CONTROLLO OPERAZIONI
DETERMINAZIONE ORBITA
DATI ORBITALI
PREVISIONI ORBITALI
UTENTI
POTENZA
INFORMAZIONI SULLO STATO OPERATIVO
COMANDI E CONTROLLO
DATI
INSEGUIMENTO

Rappresentazione schematica del sistema di controllo e di raccolta dei dati dai satelliti della serie ERTS e flusso delle informazioni raccolte dalle stazioni a terra. I dati, elaborati dal centro NASA, sono trasmessi al servizio dati che li fornisce agli utenti ed eventualmente ad altri.

fiche ed è desiderabile per quelle geologiche e per quegli obiettivi specifici nei campi dell'agricoltura, idrologia, oceanografia che riguardano le diverse ricognizioni.

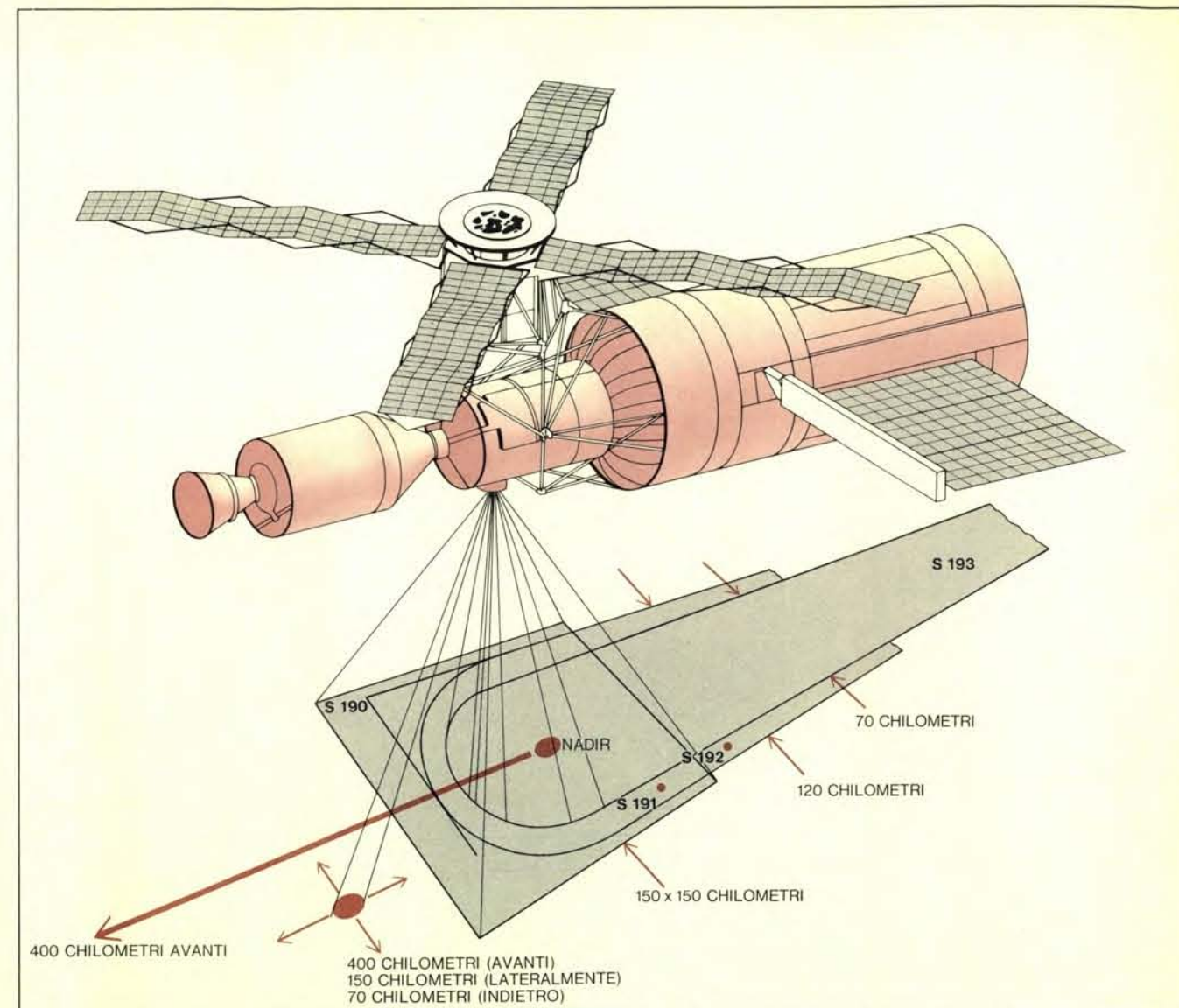
Il confronto requisiti-prestazioni per radiometri ottici è illustrato nelle figure a pagina 86. La copertura che possono dare i sistemi radiometrici attuali è molto piú grande di quella richiesta dalla maggior parte delle applicazioni, in quanto essi erano stati progettati e sviluppati primariamente per impieghi nel campo delle scienze atmosferiche che richiedono le piú ampie coperture. Per gli stessi motivi, le risoluzioni offerte non sono adeguate, tranne che per le applicazioni nel campo delle scienze atmosferiche, e le prestazioni cominciano ad avvicinarsi ai requisiti richiesti solo per i sistemi piú spinti (il radiometro 3 C a tre canali). Questi sistemi non sono a funzionamento automatico e vanno posti su orbite basse. Per questi due motivi l'impiego di questi tipi di sensori nell'immediato futuro è limitato a piattaforme spaziali con equipaggio umano. A identiche conclusioni si giunge per i radiometri a microonde.

Da questa rapida scorsa emergono deficienze e carenze dei sistemi sensori attualmente operativi e la necessità di programmi di sviluppo per migliorare le prestazioni. È già possibile anticipare quali saranno le migliorie previste per la generazione di sensori che potrà diventare operativa nel secondo quinquennio degli anni settanta e dare cosí un succinto quadro dei presumibili progressi nell'immediato futuro. Per i sistemi fotografici sarà migliorata la « qualità » delle lenti e la risoluzione dei film sarà approssimativamente raddoppiata. Analoghi progressi saranno registrati nei sistemi televisivi i quali potranno fornire, anche da quote di 800 chilometri, risoluzioni dell'ordine di 20-30 metri con coperture dello stesso ordine di quelle attuali. Essi, peraltro, non saranno ancora in grado di assicurare accuratezza metrica. Tutto sommato, pertanto, le macchine fotografiche continueranno a essere quelle che piú da vicino soddisfano i requisiti richiesti dalle diverse applicazioni. Con radiometri ottici a bande multiple si potranno ottenere, da quota 800 chilometri, risoluzioni dell'ordine di 250-300 metri, ma sembra improbabile che si possano raggiungere quelle risoluzioni (dell'ordine di 30 metri) che, come visto, sono richieste nella maggior parte delle applicazioni. La radiometria a microonde continuerà a essere utilizzabile essenzialmente solo da orbite basse. Lo stesso dicasi per i radar ad apertura sintetica (che potranno fornire risoluzioni dell'ordine di 20 metri da orbite a 400 chilometri) il cui impiego, peraltro, presupporrà la disponibilità di appropriati sistemi di conversione di energia capaci di fornire potenze dell'ordine di 10 chilowatt.

Le considerazioni sinora fatte non esauriscono la problematica della scelta dei sensori e della caratterizzazione delle missioni in funzione degli obiettivi. Parecchi altri parametri vanno infatti considerati (primo tra tutti la richiesta frequenza delle osservazioni che, già per gli obiettivi elencati nella tabella a pagina 81, varia da piú volte al giorno a una volta ogni cinque anni) ma anche una descrizione sommaria di questi parametri, delle loro mutue influenze e delle loro interazioni con gli altri elementi del sistema globale ci porterebbe a eccedere i limiti di questo discorso.

Esempi di sistemi attualmente operativi

Il 23 luglio 1972, con un po' di ritardo sulla data prevista, è stato lanciato, con successo, il primo satellite specificamente concepito per lo studio delle risorse terrestri: l'ERTS 1 (Earth Resources Technology Satellite). Il secondo, l'ERTS 2, dovrebbe essere lanciato verso la fine del 1973. Come indica lo stesso nome, il programma ERTS è un programma di sviluppo tecnologico tendente a determinare l'utilità e l'utilizzabilità dei dati da esso raccolti per lo studio delle risorse terrestri e dei problemi connessi con la vita sulla Terra. L'orbita del satellite è stata scelta in modo da consentire coperture sistematiche e ripetute, in condizioni di osservazione, di quasi tutta la superficie terrestre. L'orbita è pertanto quasi polare, circolare con un raggio di 915 chilometri, e sincrona con il Sole. Il satellite gira intorno alla Terra 14 volte al giorno (il periodo dell'orbita è di 103 minuti) e « spazza » l'intera superficie terrestre ogni 18 giorni. L'orbita è periodicamente « aggiustata » in modo che la « traccia » al suolo del satellite ripeta la copertura della superficie terrestre, alla stessa ora locale ogni 18 giorni, con una escursione massima di 37 chilometri. Il satellite è stato progettato per una vita utile di un anno in modo da poter investigare fenomeni stagionali.

Il satellite compie due funzioni: 1) ottiene direttamente dati in diverse

In questo disegno è rappresentato il laboratorio spaziale « Skylab » al quale saranno affidati i piú importanti compiti di rilevamento a distanza. Nella figura sono anche rappresentate le diverse zone che possono essere coperte dai diversi sistemi sensori installati a bordo della stazione spaziale; ciascuna zona è individuata mediante la sigla del sensore e mediante le sue dimensioni.

bande dello spettro visivo; 2) raccoglie dati trasmessi da centinaia di diverse stazioni di sensori, situate sulla superficie terrestre, e li ritrasmette a delle stazioni riceventi, terrestri, della NASA. Per la prima funzione è dotato di due sistemi sensori.

Il sistema sensorio a tubo d'immagini (*return beam vidicon*, RBV) consiste di tre camere indipendenti, operanti simultaneamente e coprenti ciascuna una sua banda spettrale nel visibile e infrarosso vicino in un arco compreso tra 0,48 e 0,83 micron. Questo sistema, che può funzionare soltanto di giorno, ha una copertura al suolo di 185 × 185 chilometri e fornisce 180 immagini al giorno. L'immagine formata sulla superficie fotosensibile è analizzata da un fascio elettronico, che produce un segnale video e impiega circa 3,5 secondi per analizzare ciascuna delle tre immagini spettrali.

Il secondo sistema sensorio (*multispectral scanner*, MSS) è costituito da una serie di rivelatori agenti simultaneamente, sensibili in quattro bande spettrali comprese tra 0,5 e 1,1 micron. Uno specchio oscillante analizza simultaneamente, con continuità e in direzione perpendicolare al moto del satellite, sei linee per ciascuna delle quattro bande spettrali; la progressione di queste sei linee lungo la direzione del moto è assicurata dal moto stesso del satellite. Le uscite dei rivelatori sono campionate, codificate e trasformate in un flusso continuo di dati di 15 megabit al secondo. Al suolo questi dati sono trasformati in immagini ricoprenti un'area approssimativamente eguale a quella delle immagini del sistema RBV con una sovrapposizione del 10 % tra due immagini consecutive.

Entrambi i sistemi operano in genere simultaneamente sopra le stesse zone durante le ore diurne. Quando il satellite è « visto » da una delle stazioni riceventi, i dati sono trasmessi al suolo in tempo reale e ivi registrati su nastro magnetico. Negli altri casi, i dati sono registrati, a bordo del satellite stesso, su due registratori a nastri video a banda larga (WBVTR, *wide band video tape recorders*) ciascuno dei quali registra e riproduce, a comando, i dati dell'uno o dell'altro sistema, con una capacità di registrazione di 30 minuti. Il sistema raccolta dati (DCS, *data collection systems*) raccoglie i dati trasmessi automaticamente da piattaforme situate sulla superficie terrestre e contenenti sino a otto sensori diver-

si, e li ritrasmette alle stazioni riceventi ogniqualvolta l'ERTS « vede » sia le piattaforme sia le stazioni riceventi dalle quali, dopo ventiquattro ore, i dati sono a disposizione degli investigatori che hanno installato le diverse piattaforme.

Le componenti essenziali del sistema manipolazione dati e controllo sono indicate nella figura a pagina 94. Il sistema consiste di due connessioni telemetriche (una per i comandi e l'altra per i dati), tre stazioni a terra, il centro controllo delle operazioni (che è il Goddard Space Flight Control), l'unità per il trattamento dei dati della NASA e l'unità servizio dati. Al centro di controllo, che opera ventiquattrore al giorno, sono effettuate tutte le operazioni necessarie a determinare se o meno il satellite e i sistemi sensori funzionano nel modo richiesto, a prendere e rendere esecutive le opportune decisioni in casi di emergenza e a pre-trattare i dati ricevuti dal sistema raccolta dati.

Il sistema manipolazione dati identifica, localizza e annota le immagini e opera su di esse tre tipi di trattamenti: normale, di precisione e speciale. Il trattamento normale è quello cui tutti i dati dei nastri video sono sottoposti e potrebbe, piú specificamente, essere considerato come un pre-trattamento. I dati dei nastri video vengono innanzitutto convertiti, a mezzo di un registratore a fascio elettronico, in immagini da 55 millimetri su film da 70 millimetri; durante questa conversione vengono anche registrate le annotazioni (alfanumeriche), la localizzazione dell'immagine e una scala di toni grigi per la taratura. Sulle immagini vengono altresí apportate le prime correzioni radiometriche e geometriche. Queste immagini su film vengono sviluppate e ispezionate per un controllo di qualità e di copertura di nubi e successivamente ingrandite (se richiesto), stampate, sottoposte a ulteriore ispezione di qualità, catalogate e, infine, distribuite. Il trattamento di precisione, effettuato solo quando specificamente richiesto dall'investigatore, è eseguito su una percentuale di dati (5 % circa delle immagini) per fornire un prodotto di elevata qualità in un formato immediatamente utilizzabile. Questi prodotti piú precisi sono necessari per quei tipi di analisi che richiedono la sovrapposizione accurata di immagini a lunghezze d'onda diverse della stessa scena. Essi sono, però, anche indispensabili per verificare le condizioni di funzionamento del satellite e individuare gli eventuali comandi correttivi da inviare al satellite. Le immagini su film da 70 millimetri subiscono un trattamento che produce immagini nel formato 240 millimetri, rimuove gli errori geometrici non corretti nel trattamento normale, localizza l'immagine con maggiore precisione, effettua la proiezione ortografica dell'immagine corretta. Il trattamento speciale è effettuato, sempre su specifica richiesta, su un selezionato numero di immagini. I dati risultanti da uno dei due trattamenti precedenti vengono analizzati, tarati, corretti e convertiti in dati registrati su nastri compatibili per trattamento su elaboratori elettronici. I dati prodotti dal sistema raccolta dati sono distribuiti agli investigatori, entro le ventiquattro ore, sotto forma di nastri magnetici, tabulati di calcolatore o schede perforate. Le fotografie di questo articolo rappresentano esempi indicativi delle immagini fornite dall'ERTS 1 (si veda l'articolo *Applicazioni delle tecniche di rilevamento a distanza in Italia* di R. Cassinis, C. M. Marino e A. M. Tonelli in « Le Scienze », n. 54, febbraio 1973).

I risultati preliminari delle analisi dei dati forniti dall'ERTS 1 sono stati presentati e discussi per la prima volta, nel corso di una tavola rotonda organizzata da William Nordberg direttore scientifico del programma ERTS, in occasione del XXIII Congresso della federazione internazionale di astronautica tenutosi a Vienna nell'ottobre del 1972.

Lo Skylab e le risorse terrestri

Il programma Skylab metterà in orbita una stazione spaziale nella quale tre differenti squadre di tecnici e scienziati effettueranno, la prima per 28 giorni e le altre due per 56 giorni, una serie di esperimenti scientifici e tecnologici, aventi per fine ultimo lo studio della Terra, del Sole, dell'uomo e della tecnologia spaziale.

Il sistema Skylab (*si veda la figura a pag. 95*) consiste essenzialmente di un laboratorio orbitale derivato dal quarto stadio del vettore Saturno opportunamente adattato per garantire l'abitabilità per lunghi periodi e contenente parte delle attrezzature necessarie per i vari esperimenti; il « Multiple Docking Assembly », connesso al laboratorio orbitale tramite l'« Airlock Module », e costituente, con i suoi due boccaporti – uno radiale e l'altro assiale – l'interfaccia con il modulo di servizio e comando che serve, tra l'altro, a trasportare lo equipaggio e la strumentazione dalla Terra al laboratorio e viceversa; l'osservatorio solare. Oltre a esperimenti di tipo medico, scientifico, ingegneristico e tecnologico, gli equipaggi dello Skylab effettueranno esperimenti rivolti allo studio delle risorse terrestri nei campi dell'agricoltura e foreste, oceanografia, idrologia, geologia e geografia. L'interesse di questi esperimenti risiede, tra l'altro, nel ruolo giocato dall'uomo. Si prevede che lo Skylab sia equipaggiato con i seguenti sistemi di sensori. 1) S 190. Si tratta di un sistema fotografico a 6 bande – quattro per film in bianco e nero, una per film a colori e un'altra per film infrarosso a colori – con distanza focale di 15 centimetri e risoluzione di 300 metri dalla distanza di 400 chilometri. L'equipaggio provvederà a caricare i film, cambiare i filtri, controllare i tempi e gli intervalli di esposizione e, infine, a tutte le necessarie operazioni di manutenzione e eventuale sostituzione di parti non funzionanti. 2) S 191. È uno spettrometro infrarosso a due bande con rilevatore termico raffreddato da un sistema a ciclo chiuso, miniaturizzato. I dati sono registrati a bordo su nastro magnetico, che viene decommutato al suolo. Il contributo dell'astronauta è essenziale in quanto egli deve acquisire visualmente (con l'aiuto di un telescopio) aree di diametro dell'ordine di 2 chilometri e attivare la sequenza di taratura automatica prima e dopo ciascuna rivoluzione. 3) S 192. È un radiometro a scansione ottica con 13 bande nel visibile e nel vicino e lontano infrarosso con una risoluzione prevista di circa 80 metri. L'astronauta provvederà a calibrare lo strumento, a cambiare i nastri magnetici e a effettuare le altre operazioni di manutenzione e sostituzione. 4) S 193. È un sistema combinato attivo e passivo comprendente un radiometro, uno scatterometro e un altimetro con antenna parabolica ricevente e trasmittente radiazioni con duplice polarizzazione (i dati sono registrati su nastro magnetico). Sarà usato essenzialmente per valutare, simultaneamente, la sezione d'urto di diffusione radar e la emissività del terreno e del mare e per confrontare misure di brillanza a due diverse frequenze (mediante correlazione con i dati del sistema S 194). L'astronauta dovrà scegliere i modi di scansione (contigua o non contigua nel senso del moto od ortogonalmente a esso) e i modi di operazione (radiometro, scatterometro, radiometro-scatterometro, altimetro). 5) S 194. È un radiometro a microonde operante alla frequenza di 1,41 GHz con ampiezza di banda di 27 MHz. I dati sono registrati su nastro magnetico. L'astronauta mette in azione lo strumento e attiva la sequenza di calibratura.

Alla fine di ciascuna delle tre missioni saranno riportati al suolo 4800 fotogrammi di film, in bianco e nero, da 70 millimetri; 2400 fotogrammi di film a colore, sempre da 70 millimetri, e quattro bobine di nastro magnetico (ciascuna di 2000 metri) da 2,5 centimetri con 28 piste e con 800 bit al centimetro per pista. Non è prevista, per ora, la teletrasmissione, né in tempo reale né durante la missione, dei dati raccolti dallo Skylab. Tutti i dati saranno registrati su film o su nastri: questi ultimi saranno convertiti dal formato originale (nastro analogico a 28 piste) a quello finale (nastro a 14 piste); gli originali sono archiviati e i duplicati verranno inoltrati agli opportuni centri per il loro trattamento.

Considerazioni finali

In questa succinta presentazione non si sono ovviamente potuti discutere tutti gli aspetti del problema. Alcuni sottosistemi – quale, per esempio, quello connesso con il trattamento dei dati di cui è data una rappresentazione a blocchi nella figura in questa pagina – per la loro complessità e rilevanza meriterebbero una trattazione a sé stante. Non si può comunque fare a meno di sottolineare e commentare un fatto di estrema importanza: le capacità attuali della tecnologia già superano, di gran lunga, la capacità di analizzare in modo efficiente e di utilizzare i copiosi dati da esse fornite. Sarà possibile potenziare queste capacità solo se si organizzano gruppi interdisciplinari che, potendo valersi di abilità, conoscenze e bagagli culturali diversi, saranno in grado di individuare ed « estrarre » le informazioni che detti dati contengono.

Questo schema a blocchi mostra l'intero ciclo cui vengono sottoposti i dati ottenuti dai sistemi telesensori. La fase di vagliatura porta alla selezione dei dati interessanti dopo un'accurata verifica al suolo, verifica che mantiene inalterata tutta la sua importanza.

Si consideri, per esempio, il problema della « discriminazione » inteso come processo consistente nella successiva classificazione di larghe classi in subclassi sempre piú piccole e piú dettagliatamente definite. In questo processo si possono distinguere diverse fasi – differenziazione, identificazione, caratterizzazione – sempre piú raffinate, che richiedono tecniche e metodologie via via piú sofisticate. Il problema è: data la « firma » elettromagnetica del materiale in esame (data cioè una certa collezione di proprietà osservate) estrarre da essa le informazioni necessarie per procedere alla suddetta « discriminazione ». Questo è un campo di ricerca quasi vergine. Per quanto, invero, si possa già pensare a quattro metodologie diverse di discriminazione – spettrale, spaziale, per polarizzazione e temporale –, solo la prima è abbastanza ben definita e frequentemente adoperata (anche se si è ben lungi dall'averne esaurite tutte le possibilità).

Al livello piú elementare la discriminazione spettrale è a tutti familiare: essa fa uso dei colori dello spettro visibile osservati direttamente dall'occhio umano o tramite fotografie, in bianco e nero o a colori. La stessa metodologia può essere applicata, in linea di principio, in altre bande dello spettro elettromagnetico (singolarmente o, simultaneamente, con la tecnica multispettrale). Esistono però problematiche che hanno ancora bisogno di ulteriori studi quali, per esempio, quelle afferenti alle metodologie di calibrazione e interpretazione dei dati (preferibilmente a mezzo di processi di elaborazione automatica) che devono tener conto del carattere non deterministico delle firme. Con la discriminazione spaziale si tende a determinare forma e struttura dell'elemento osservato. In questo campo non è ancora nemmeno ben assodato quali siano le « caratteristiche da misurare » e come esse si presentino. Con la discriminazione per polarizzazione si tenta di sfruttare le differenze nell'orientamento del vettore campo elettrico risultante delle radiazioni elettromagnetiche riflesse e/o emesse dall'oggetto. Questa sembra essere una tecnica promettente per la discriminazione dei tipi di terreno. Con la discriminazione temporale si cerca essenzialmente di sfruttare le caratteristiche delle variazioni diurne e stagionali cui, per esempio, sono affette tutte le piante e i terreni. Anche in questo caso, però, si sa ben poco degli effetti di queste variazioni temporali sulle porzioni dello spettro elettromagnetico diverse dal visivo. In questo contesto va sottolineata l'importanza dello studio della « storia termica » di un dato oggetto ai fini di una sua caratterizzazione.

Per concludere, quindi, la potenzialità di impiego di satelliti artificiali per lo studio delle risorse terrestri sono molteplici e tutte di grande interesse. Esistono però ancora notevoli problematiche, tutte di tipo inter- e/o multidisciplinare, che vanno concretamente affrontate e risolte prima che queste potenzialità possano essere tradotte in attualità e dare origini a servizi utili alla comunità umana. È pertanto auspicabile che vengano, ovunque e al piú presto, promossi, organizzati, coordinati e potenziati gli opportuni programmi interdisciplinari di ricerche e sviluppo, non solo su piano internazionale ma anche, e soprattutto, sui piani nazionali sui quali, in ultima analisi, va affrontata la sfida presentata dall'analisi di enormi masse di dati e dalla estrazione di informazioni *praticamente utilizzabili*.

La rotazione nell'astrofisica delle alte energie

Quale è la fonte di energia dei pulsar, dei quasar e degli altri oggetti strani? Può essere l'energia gravitazionale che si converte in energia rotazionale quando un oggetto grande si contrae in uno piccolo

di Franco Pacini e Martin J. Rees

L'universo rivelato dall'astronomia negli ultimi tre secoli era un universo calmo all'infuori di qualche sporadica esplosione di nova. Questo quadro è andato a pezzi con la scoperta di fenomeni che coinvolgono variazioni rapide e violente. Esempio tipico nella nostra galassia è il comportamento della Nebulosa del Granchio nella costellazione del Toro, costituita dagli avanzi in espansione dell'esplosione di una supernova dell'anno 1054 descritta dagli astronomi cinesi. Questo oggetto emette a tutte le lunghezze d'onda, dalle onde radio ai raggi X, e l'emissione totale è equivalente a 100 000 volte l'energia del Sole. La radiazione è prodotta soprattutto per meccanismo di sincrotrone, cioè da elettroni veloci che si muovono a spirale verso l'esterno lungo le linee di forza del campo magnetico della nebulosa. All'esterno della nostra galassia vi è un numero maggiore di spettacolari fenomeni esplosivi comprendenti alcuni meccanismi di radiazione che non sono legati alle stelle. Certi oggetti extragalattici hanno inoltre caratteristiche assai simili a quelle della Nebulosa del Granchio: per esempio la galassia NGC 1275 ha una struttura a filamenti. Sembra che tutti questi fenomeni abbiano gli stessi elementi fondamentali: particelle veloci e forti campi magnetici.

La stella centrale della Nebulosa del Granchio è un pulsar: un oggetto che emette radiazione a rapidi impulsi. Sembra che valga la pena approfondire l'idea che il pulsar della Nebulosa del Granchio (e i pulsar in generale) possa essere il prototipo di tutti i fenomeni di astrofisica delle alte energie e che ci permetta perciò di studiare, a relativa portata di mano entro la nostra galassia, gli stessi meccanismi che operano su grande scala nei quasar e nelle radiosorgenti extragalattiche.

Una stella normale dipende dall'energia liberata quando i nuclei di idrogeno (i protoni) formano per fusione nuclei piú pesanti nella parte piú interna della stella. Esiste un limite superiore all'efficienza dei combustibili nucleari imposto dalla natura delle forze che tengono insieme le particelle nucleari. L'efficienza massima è dello 0,8 per cento e corrisponde alla massima energia che lega una particella a un nucleo atomico. La maggior parte dell'energia che si produce in una stella proviene dalla trasformazione di idrogeno in nuclei di elio in condizioni per cui ogni nucleo di idrogeno ha un'energia di solo poche migliaia di elettronvolt. È difficile vedere come queste energie relativamente modeste possano dare origine a elettroni veloci che generano radiazione di sincrotrone e che si muovono con velocità prossime a quella della luce. Sembra sia necessaria un'altra fonte di energia e tale fonte potrebbe essere la gravità.

Nelle stelle di tipo solare l'energia gravitazionale ha una importanza relativa dato che l'energia gravitazionale che il Sole avrebbe liberato dalla sua nascita, quando si è contratto da una nube di gas fino alle dimensioni attuali, non sarebbe stata sufficiente a mantenere la luminosità attuale per molto piú di 10 milioni di anni. L'energia nucleare del Sole, d'altra parte, può mantenere allo stato attuale la nostra stella per vari miliardi di anni. In altre parole l'energia gravitazionale che lega un singolo protone al resto della stella è molto inferiore all'energia nucleare che lega il protone al nucleo di elio. Perciò l'energia nucleare liberata da una stella normale contribuisce solo in piccola misura al bilancio energetico.

L'energia gravitazionale di legame di una stella può però aumentare in due modi. In primo luogo cresce se si aggiungono piú particelle; aumentando la massa del sistema aumenta la forza gravitazionale che è direttamente proporzionale alla massa. In secondo luogo cresce se la stella si contrae: la forza attrattiva tra le particelle è inversamente proporzionale alla distanza che le separa e l'energia di legame dipende dall'inverso di tale distanza; pertanto, se le particelle vengono compresse in modo da essere due volte piú vicine, la stella ha un'energia gravitazionale di legame doppia rispetto a quella che aveva prima.

Per una stella con la massa del Sole l'energia gravitazionale di legame supererebbe l'energia nucleare di legame se la stella si contraesse fino a un raggio inferiore a 100 chilometri. Se il Sole terminasse la sua vita come stella di neutroni, con un raggio di qualche decina di chilometri, durante la sua evoluzione avrebbe liberato piú energia gravitazionale di quella ottenibile dalle reazioni nucleari. Metà di questa energia è disponibile sotto forma di rotazione. Perciò una stella « morta » inizia una nuova vita con piú energia a sua disposizione di quanta ne abbia mai avuta prima! L'energia gravitazionale può anche superare l'energia nucleare nelle stelle supermassicce: ipotetici oggetti piú massicci di milioni di Soli che possono avere importanza nei nuclei attivi di certe galassie.

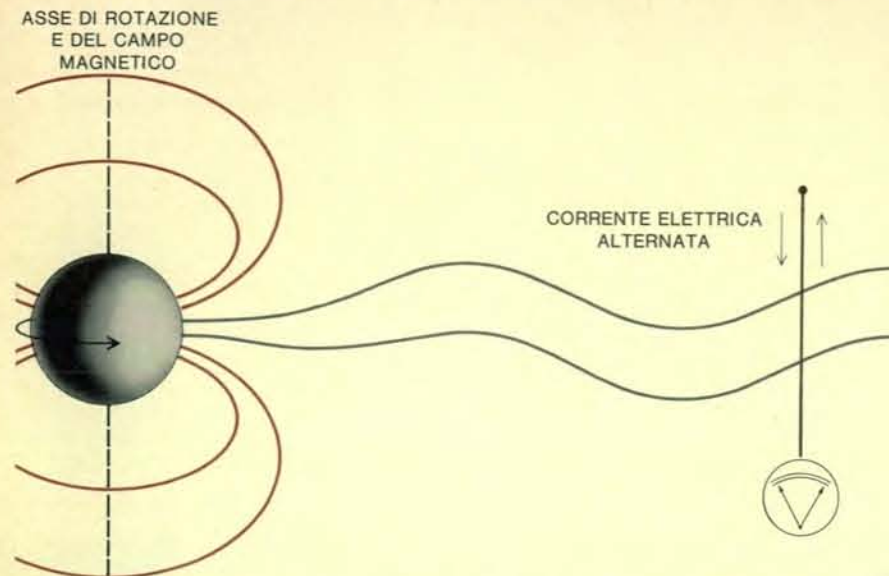
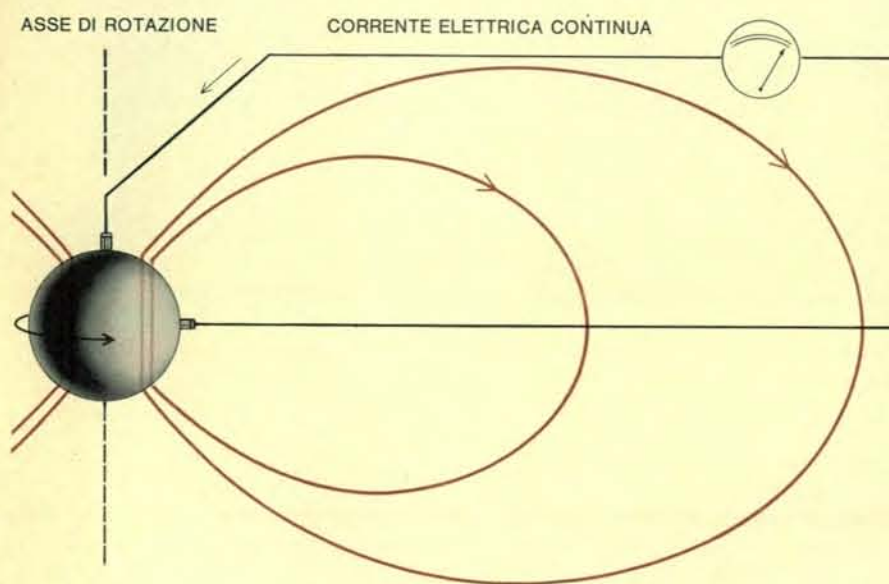
Due caratteristiche generali del collasso gravitazionale rivestono un'importanza fondamentale nell'astrofisica delle alte energie. Innanzitutto, se il sistema iniziale è in rotazione, quando esso si contrae il campo gravitazionale si oppone alle forze centrifughe. Siccome si deve conservare il momento della quantità di moto, l'energia gravitazionale si trasforma gradualmente in energia cinetica di rotazione. In secondo luogo, se vi sono correnti elettriche interne che generano un campo magnetico, l'energia associata con queste correnti cresce quando sono compresse e tale energia deve essere fornita dalla gravitazione.

La Nebulosa del Granchio (*sopra*) nella costellazione del Toro è composta dai resti ricchi di filamenti di una supernova esplosa nella nostra galassia nel 1054. Gli elettroni che si muovono a spirale a velocità prossima a quella della luce entro il campo magnetico della nebulosa emettono radiazione di sincrotrone a tutte le lunghezze d'onda, dalla regione radio ai raggi X.

La galassia NGC 1275 (*sotto*) ha una struttura a filamenti molto simile a quella della Nebulosa del Granchio e anch'essa emette radiazione di sincrotrone da elettroni in moto a spirale attraverso il suo campo magnetico. NGC 1275 è uno dei numerosi oggetti esterni alla Galassia che si comportano in modo assai simile alla Nebulosa del Granchio, ma su scala molto maggiore.

Per esempio, il Sole ruota circa una volta al mese e ha un campo magnetico globale alla superficie di circa un gauss (all'incirca uguale al campo magnetico terrestre). Se il Sole fosse compresso fino alle dimensioni di una stella di neutroni compirebbe circa 1000 rivoluzioni per secondo e acquisterebbe un campo magnetico di 10 miliardi (10^{10}) di gauss.

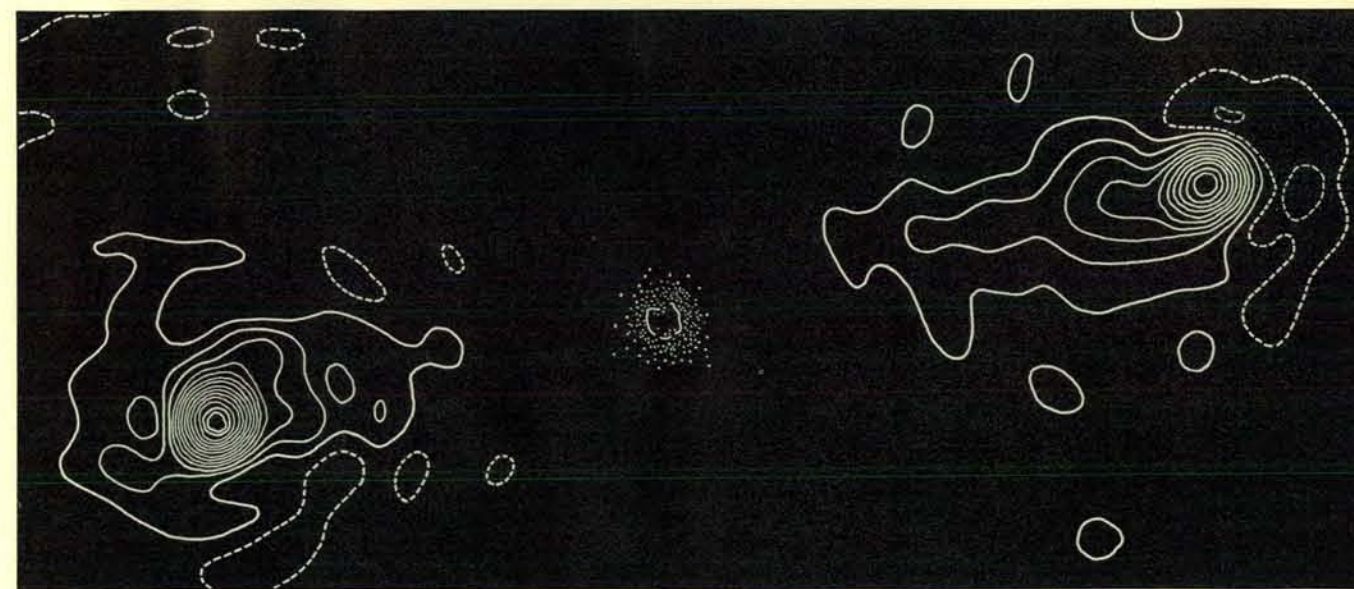
È proprio la combinazione di una rotazione rapida e degli enormi campi magnetici che risultano dalla contrazione gravitazionale che oggi si ritiene abbia un ruolo fondamentale nei fenomeni di astrofisica delle alte energie. La migliore dimostrazione proviene dalla scoperta dei pulsar e dall'ipotesi che siano stelle di neutroni in rapida rotazione. Questa ipotesi, di cui il primo sostenitore fu Thomas Gold della Cornell University, è stata poi accettata generalmente. Il pulsar che ruota piú rapidamente, 30 volte al secondo, è quello al centro della Nebulosa del Granchio. La sua posizione è in accordo con la teoria che le stelle di neutroni si formino nell'esplosione delle supernove. Il tempo di vita degli elettroni energetici della Nebulosa del Granchio, responsabili della radiazione di sincrotrone emessa alle lunghezze d'onda ottiche e dei raggi X, è inferiore a 1000 anni; queste particelle non possono perciò essere coeve all'esplosione iniziale che ha creato la nebulosa. Nel 1958, L. J. Woltjer, un astronomo olandese la cui tesi è diventata un testo classico sulla Nebulosa del Granchio, ha messo in evidenza come entro la nebulosa ci debba essere una qualche sorgente continua di particelle veloci. L'importanza principale del pulsar della Nebulosa del Granchio sta nel modo in cui risolve questo vecchio problema di come avviene il rifornimento di elettroni energetici. L'intervallo in cui variano la massa e il raggio delle stelle di neutroni è stato definito abbastanza bene dagli studi teorici. Se è noto a quale velocità ruota una stella e quanto in fretta rallenta, è relativamente facile calcolare con quale velocità perde energia. Il periodo del pulsar della Nebulosa del Granchio cresce di circa 15 microsecondi all'anno e ciò significa che il pulsar perde energia 100 000 volte piú in fretta del Sole. Questo ritmo si accorda particolarmente bene con la energia richiesta per mantenere l'emissione di sincrotrone e ciò vuol dire che ci deve essere qualche meccanismo che converte l'energia di rotazione del pulsar in particelle veloci con un'efficienza che è certamente superiore al miglior valore dell'1 per cento circa raggiunto con gli acceleratori in laboratorio.

Quale processo fornisce la coppia che può estrarre energia da una stella di neutroni in rotazione? Non ci sono dubbi che il campo magnetico è assai importante. Un magnete che gira nel vuoto irraggerà onde elettromagnetiche con una frequenza pari alla sua frequenza di rotazione ammesso che l'asse del campo magnetico non coincida esattamente con l'asse di rotazione. Le onde trasportano energia con un tasso che dipende dall'intensità del campo

Un campo obliquo dipolare è un meccanismo che può estrarre energia da una stella di neutroni in rotazione o da un pulsar e farlo rallentare gradualmente. L'asse del campo magnetico stellare *(in colore)* non coincide con l'asse di rotazione, ma forma con esso un certo angolo. Quando la stella magnetizzata gira, irradia onde elettromagnetiche *(in grigio)* che hanno frequenza uguale alla frequenza di rotazione, ammesso che la stella sia in un vuoto sufficiente e non sia immersa in gas troppo ionizzato. Le onde accelerano gli elettroni a velocità prossime a quella della luce e li rendono capaci di emettere radiazione di sincrotrone. La quantità di energia che viene estratta dipende dall'intensità del campo magnetico stellare e dalla sua velocità di rotazione.

L'induttore unipolare è un altro meccanismo in grado di estrarre energia da una stella di neutroni rotante o da un pulsar. Esso può funzionare anche se la stella è circondata da troppo plasma (gas ionizzato) perché possano propagarsi onde elettromagnetiche pure. Se in laboratorio si fa ruotare una sfera conduttrice magnetizzata e se si collega un terminale di un circuito esterno a un punto dell'asse e l'altro terminale a una spazzola fissa a contatto con l'equatore della sfera, il galvanometro inserito nel circuito segnalerà un passaggio di corrente e la sfera rallenterà. Nel caso di una stella di neutroni il plasma che la circonda ha la funzione del filo del circuito esterno. Anche in questo caso l'energia proviene dall'energia di rotazione della stella e la coppia frenante cui è sottoposta la stella è all'incirca la stessa di quella che si avrebbe se l'energia fosse portata via dalle onde elettromagnetiche. La differenza principale si ha nel fatto che l'induttore unipolare, a differenza del campo obliquo dipolare, funziona, anche nel caso di una stella i cui poli magnetici si trovino sull'asse di rotazione.

La radiosorgente Cygnus A ha l'aspetto di un oggetto debole, visibile nel centro di questa fotografia presa col telescopio riflettore di 5 metri di Monte Palomar. Si pensava all'inizio che l'intensa radiazione di sincrotrone potesse essere emessa da due galassie in collisione. Si ritiene oggi che la radiazione si origini per qualche tipo di attività intrinseca di una singola galassia.

magnetico e dalla velocità di rotazione. Se il pulsar della Nebulosa del Granchio avesse un campo il cui asse non coincidesse con l'asse di rotazione, l'intensità del campo alla superficie della stella dovrebbe essere circa 10^{12} gauss per produrre il rallentamento osservato nel periodo del pulsar *(si veda la figura in alto a fronte)*.

Sebbene un campo di questo tipo sia fenomenologicamente intenso rispetto ai campioni di laboratorio, cade tuttavia nell'intervallo che ci si aspetterebbe se il campo magnetico di una stella normale fosse « congelato » e amplificato quando la stella è compressa in una stella di neutroni. Mediante calcoli precisi, James E. Gunn e Jeremiah P. Ostriker dell'Università di Princeton e altri hanno dimostrato che le onde elettromagnetiche di bassissima frequenza emesse da un pulsar in rotazione possono avere un'efficienza piú che sufficiente per accelerare particelle cariche. Qualsiasi particella iniettata viene immediatamente accelerata e verrebbe trascinata via sulla cresta dell'onda, se vi si trovasse, a una velocità prossima a quella della luce (si veda l'articolo *La natura dei pulsar* di Jeremiah P. Ostriker, in « Le Scienze », n. 32, aprile 1971). È ancora un argomento dibattuto se si può considerare o meno la zona intorno al pulsar della Nebulosa del Granchio sufficientemente prossima al vuoto per permettere la presenza di queste onde elettromagnetiche di bassa frequenza.

Anche se la stella fosse circondata da troppo plasma (gas ionizzato), perché si possano propagare onde elettromagnetiche, pure può ancora perdere energia per l'equivalente cosmico dell'induttore unipolare, meccanismo studiato la prima volta da Michael Faraday. In laboratorio un induttore unipolare è una sfera metallica magnetizza-

La mappa di Cygnus A alla lunghezza d'onda radio di sei centimetri rivela che la sorgente emette radiazione da due aree che si trovano in posizione quasi simmetrica rispetto all'oggetto centrale. Le linee di livello della mappa rappresentano le gradazioni della brillanza superficiale della radiosorgente. Questa mappa è stata fatta da S. Mitton e M. Ryle dell'Università di Cambridge.

ta. Un terminale di un circuito esterno è attaccato a uno dei poli di rotazione della sfera e l'altro terminale è una spazzola statica in contatto con l'equatore della sfera. Quando si fa ruotare la sfera, un galvanometro nel circuito registra il passaggio di una corrente elettrica. Nel caso di una stella di neutroni il plasma circostante sostituisce il filo del circuito esterno e si stabiliscono correnti elettriche. L'energia che aziona questo meccanismo proviene, come nel caso di laboratorio, dall'energia di rotazione del corpo che gira. Per un campo magnetico di data intensità la coppia fornita dall'induttore unipolare è circa la stessa che si avrebbe se l'energia fosse trasportata dalle onde elettromagnetiche: perciò il campo magnetico della stella sarebbe ancora di circa 10^{12} gauss. L'unica differenza è che l'induttore unipolare può operare in un sistema simmetrico intorno al proprio asse di rotazione. Il pulsar della Nebulosa del Granchio potrebbe generare differenze di potenziale anche di 10^{14} volt con meccanismo di induttore unipolare e le particelle potrebbero raggiungere alte energie per un processo che è essenzialmente elettrostatico come in un generatore di van de Graaff. Le forze esercitate dai campi elettrici alla superficie di una stella di neutroni supererebbero la forza di gravità di un fattore di milioni o perfino di miliardi. Come hanno dimostrato Peter M. Goldreich e William H. Julian dell'Istituto di Tecnologia della California; ciò significa che la stella può fornire un numero sufficientemente alto di particelle elettricamente cariche per mantenere un'alta conduttività nell'ambiente circostante.

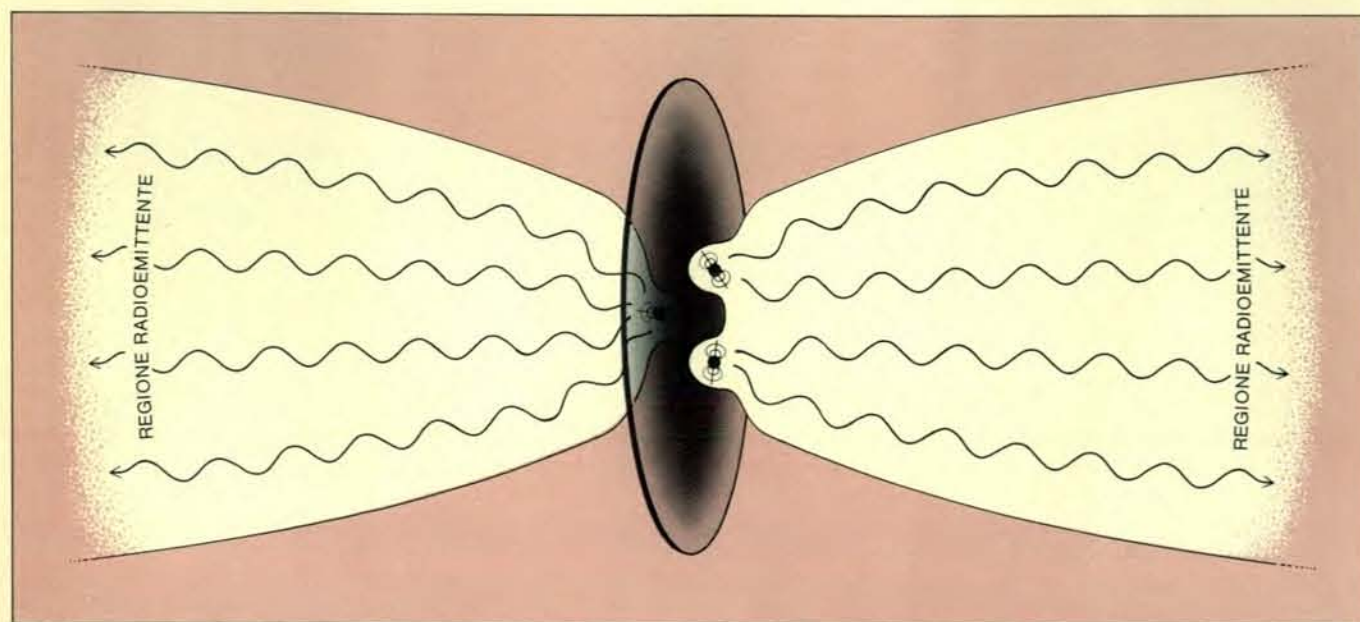
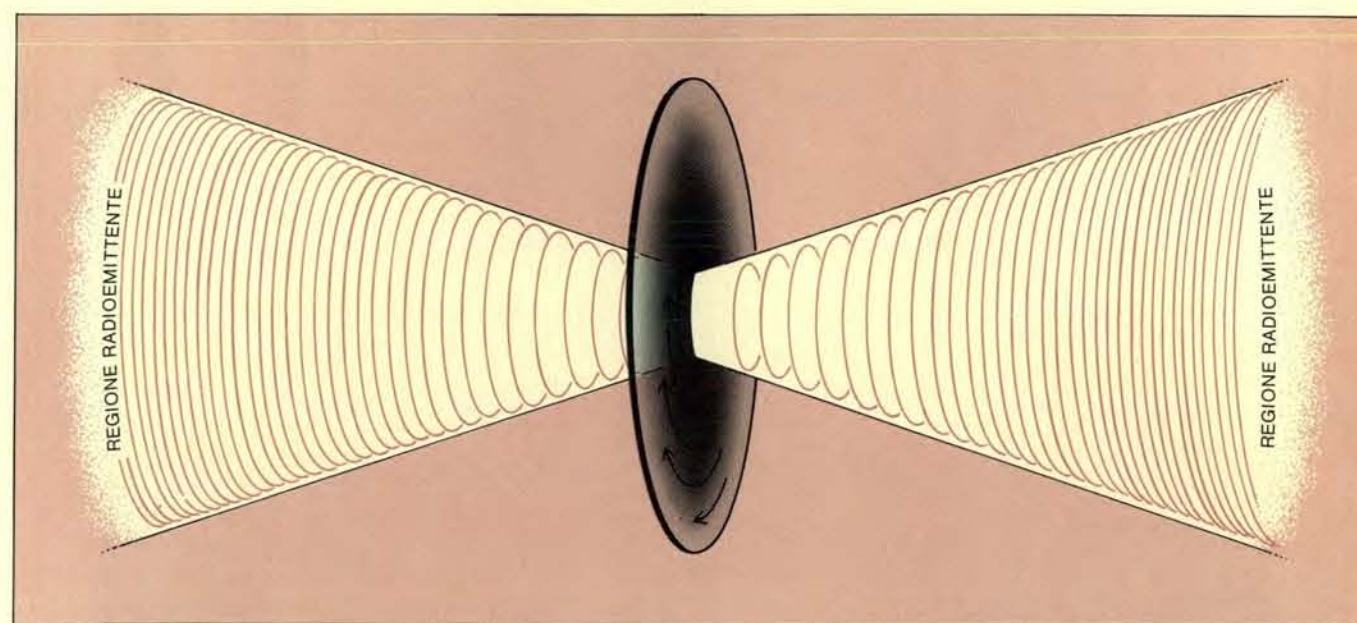
Quale sarà l'evoluzione futura di un sistema quale la Nebulosa del Granchio? In qualche decina di migliaia di anni i resti della supernova avranno subito un'espansione tale da fondersi col mezzo interstellare generale e le particelle di alta energia prodotte dal pulsar contribuiranno allo sfondo generale di raggi cosmici della Galassia. Nonostante che la rotazione del pulsar rallenti, gli impulsi potranno continuare per milioni di anni. Gli altri 60 o 70 pulsar noti si sono formati presumibilmente in eventi simili a quello della supernova del 1054. Sopravvivono poche tracce dei loro filamenti salvo per la nebulosità che circonda il pulsar nella costellazione della Vela. Il pulsar della Vela ha un periodo che è il secondo per brevità, e può perciò essere il secondo per età.

Alla fine l'emissione elettromagnetica di un pulsar potrebbe diventare così piccola che la pressione di radiazione a essa associata non potrebbe più opporre resistenza al mezzo interstellare e l'emissione radio impulsata si estinguerebbe. La stella di neutroni cesserebbe di irraggiare a meno che fosse riscaldata dalla materia interstellare che cade su di essa; materia che la costringerebbe a emettere nella regione spettrale dei raggi X. Se la stella fosse isolata nel mezzo interstellare il tasso con cui aumenterebbe la propria massa sarebbe assai basso e lo stesso avverrebbe per la sua luminosità nella regione dei raggi X. Un oggetto cosí debole sarebbe rivelabile solo entro circa 10 parsec (33 anni luce) dal Sole. Il contributo totale di tutti i pulsar morti nella Galassia (che potrebbe essere di un miliardo) produrrebbe però uno sfondo apprezzabile di raggi X. D'altra parte, se un pulsar morto fosse situato in una nube densa di materia interstellare, o fosse membro di un sistema doppio di stelle, allora il tasso con cui si aggregherebbe materia sarebbe maggiore e aumenterebbe perciò la emissione di raggi X. Si pensa di poter spiegare con questo modello molte sorgenti galattiche discrete di raggi X che hanno l'aspetto di oggetti compatti e giacciono su un'orbita stretta, quasi radente, rispetto a un compagno. L'emissione di raggi X risulta principalmente dall'impatto di gas accelerato dalla gravità a circa metà della velocità della luce sulla crosta della stella di neutroni. Una rotazione lenta può ancora produrre variazioni periodiche nelle luminosità osservate dei raggi X se per esempio il campo magnetico incanalasse la materia accumulata su certe parti della superficie stellare. Sebbene la velocità di rotazione fornisca ancora « l'orologio », è l'energia gravitazionale della materia che cade, e non la rotazione, che sta alla base delle sorgenti di raggi X.

Vi sono stati vari tentativi di usare simili ipotesi per spiegare l'attività violenta che è stata osservata negli ultimi anni nei nuclei delle galassie. La luminosità di un nucleo galattico può arrivare a 10^{13} volte la luminosità del Sole alle lunghezze d'onda visibili. La potenza totale emessa è forse anche maggiore se il nucleo irraggia fortemente nell'infrarosso che è fortemente tagliato, per un osservatore terrestre, dal vapor acqueo dell'atmosfera. Queste luminosità sono circa 100 volte maggiori della luminosità di una galassia come la nostra. Inoltre questa colossale produzione di potenza è concentrata in una piccola regione centrale; si è osservato inoltre che è variabile alle lunghezze d'onda sia ottiche sia radio con periodi di mesi o anche meno. Se le dimensioni del volume che emette la radiazione superano pochi mesi luce, queste variazioni verrebbero eliminate dato che la luce impiega un tempo finito per viaggiare da una parte all'altra della sorgente. Si ha perciò l'impressione che in queste galassie vi sia una regione non molto più grande del sistema solare in grado di liberare una potenza 100 volte maggiore di quella di una galassia normale.

Un esempio di questo tipo di oggetti sono le galassie Seyfert: galassie a spirale con nuclei assai brillanti la cui luce proviene più dal gas incandescente e dalle particelle veloci che dalle stelle. Un secondo esempio sono i quasar che molti astronomi ritengono nuclei galattici « iperattivi »; manifestano un'attività molto più alta dei nuclei Seyfert e la loro luminosità sommerge il contributo delle stelle della galassia circostante. I quasar sono spesso associati con forti ed estese radiosorgenti che presentano la caratteristica di irraggiare da due aree principali poste quasi simmetricamente rispetto all'oggetto centrale. Si ammette che gli elettroni, che emettono radiazione di sincrotrone alle lunghezze d'onda radio, abbiano in qualche modo origine nell'oggetto centrale. Queste estese radiosorgenti doppie si trovano anche associate con le galassie ellittiche, i cui nuclei peraltro non sono attivi con evidenza in altro modo. Le radiosorgenti più forti, come la radiosorgente nella costellazione del Cigno, nota come Cygnus A, furono i primi esempi di attività violenta a essere scoperti fuori della Galassia. All'inizio si avanzò la ipotesi che potessero rappresentare collisioni tra galassie, ma ora è chiaro che rappresentano un tipo di attività intrinseca di una singola galassia.

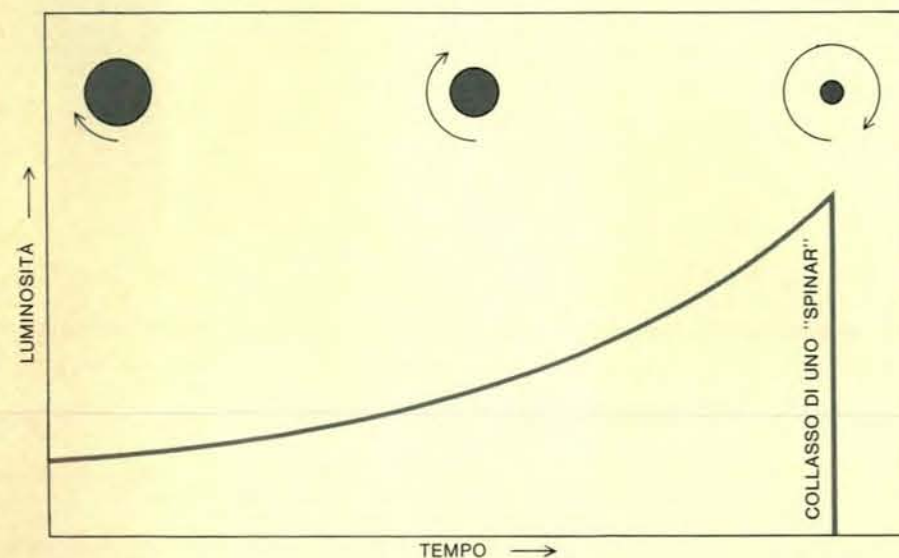
La natura di questa attività nelle radiogalassie, nei quasar e negli oggetti analoghi è ancora misteriosa; molti ritengono che possa racchiudere qualche principio fondamentalmente nuovo della fisica. Ciò aumenta l'interessante prospettiva che gli studi astronomici possano rivelare qualche nuova legge della natura la cui scoperta si sottrae alle esperienze di laboratorio. Sarebbe però avventato accettare questa conclusione senza prima indagare su tutte le possibilità racchiuse nei principi della fisica convenzionale. La migliore speranza di interpretare questi fenomeni secondo le linee convenzionali sta forse nell'applicazione delle stesse idee che hanno avuto tanto successo nel caso della Nebulosa del Granchio.

Una possibile spiegazione per le forti radiosorgenti extragalattiche è che la frequenza delle esplosioni delle supernove sia molto maggiore nei nuclei galattici attivi che nei nuclei galattici ordinari (forse perché le stelle che le compongono sono diverse) e che come risultato vengano continuamente prodotti giovani pulsar energetici. La Nebulosa del Granchio mostra che un pulsar può produrre intorno a sé una zona prossima al vuoto esercitando una enorme pressione di radiazione elettromagnetica. Qualcosa di analogo potrebbe avvenire nel gas che inviluppa i pulsar dei nuclei galattici: ogni pulsar che si forma può creare una cavità in espansione che contiene solo campi elettromagnetici ed elettroni veloci. Studi precisi sulla natura delle variazioni alle lunghezze d'onda radio suggeriscono che tali nubi di particelle e campi si espandano e scompaiano in

I numerosi pulsar originatisi da frequenti esplosioni di supernove nel nucleo di una galassia possono offrire una spiegazione delle forti radiosorgenti extragalattiche. Un pulsar di recente formazione può creare nel gas circostante una cavità in espansione (*in bianco*) che contiene solo campi elettromagnetici ed elettroni veloci. Se il gas nella parte centrale della galassia (*in grigio*) è in rotazione, tende ad appiattirsi in un disco. Le onde elettromagnetiche (*in nero*) sfuggiranno dal nucleo galattico preferenzialmente lungo l'asse di rotazione. Queste onde focalizzate in due fasci ben collimati interagiscono poi col mezzo intergalattico (*in colore*) e generano elettroni veloci che irradiano alle frequenze radio. Si possono probabilmente spiegare proprio secondo queste linee ipotetiche le simmetrie e tutte le altre svariate proprietà che sono caratteristiche delle radiosorgenti doppie.

Un corpo enorme in mezzo a una galassia, con una massa 100 milioni di volte quella del Sole, può servire come alternativa per spiegare le estese radiosorgenti doppie rispetto al modello che prevede molti pulsar. A una stella tanto supermassiccia mancherebbe la rigidità di una stella di neutroni e non ruoterebbe come un corpo solido; le parti interne tenderebbero a ruotare più velocemente di quelle esterne. Le linee di forza del suo campo magnetico (*linee colorate*) si avvolgerebbero a spirale e risulterebbero amplificate da questa rotazione differenziale. Il campo può esplodere come una molla da ambo le parti del disco. Una forte radiazione di sincrotrone alle lunghezze d'onda radio verrebbe emessa dagli elettroni veloci che si muovono nelle regioni di spazio in cui il campo risulta compresso e rinforzato a causa dell'interazione col mezzo intergalattico (*colore chiaro*).

un intervallo di pochi mesi. Si possono allora identificare le cavità con successive esplosioni nel dominio radio (*outburst*). Tali esplosioni possono allora creare e mantenere la riserva di particelle necessarie per rendere conto della luminosità della sorgente alle altre lunghezze d'onda.

Se la materia che costituisce il nucleo galattico è in rotazione, tende ad appiattirsi in un disco. Le particelle veloci e i campi magnetici tendono allora a sfuggire dal nucleo, preferenzialmente lungo l'asse di rotazione. Ciò può costituire in parte la spiegazione per la simmetria direzionale delle estese radiosorgenti doppie. Alcuni astrofisici ritengono persino che le onde elettromagnetiche di bassa frequenza generate da corpi in rotazione simili al pulsar del nucleo possano venire incanalate lungo tubi nel mezzo intergalattico, come le onde di alta frequenza sono incanalate nelle guide d'onda costruite dall'uomo, per accelerare gli elettroni nelle regioni radioemittenti.

Invece di considerare il nucleo galattico come composto di molti oggetti piccoli ciascuno della massa di una stella ordinaria, si può considerare come alternativa la possibilità che tutta la massa sia concentrata in una o poche stelle supermassiccie ciascuna di massa 10^8 (o piú) volte la massa del Sole. Il collasso gravitazionale di stelle supermassiccie fu suggerito la prima volta da William A. Fowler dell'Istituto di Tecnologia della California e da Fred Hoyle dell'Università di Cambridge per spiegare l'energia delle radiosorgenti forti. Un oggetto di questo tipo assomiglierebbe solo vagamente a una stella ordinaria. Vi sarebbe un equilibrio delicato tra la sua gravità e la pressione di radiazione interna. Sarebbe instabile a meno che possedesse un campo magnetico o fosse soggetto a rotazione, ma naturalmente è soltanto una stella supermassiccia magnetizzata in rotazione che sembrerebbe indicata se vi fosse davvero una somiglianza coi processi che avvengono nella Nebulosa del Granchio.

Un pulsar ruota piú lentamente e si indebolisce con l'età. Dopo che la stella originale è esplosa violentemente come supernova, i resti diventano una stella di neutroni e un pulsar. Alla gravità si oppongono principalmente le forze nucleari mentre la rotazione non contribuisce apprezzabilmente al suo equilibrio. I pulsar piú giovani ruotano piú velocemente e irraggiano piú intensamente. Si indeboliscono gradualmente e poi rallentano quando le coppie elettromagnetiche che agiscono all'interno trasportano via l'energia.

Uno « spinar » ruota piú velocemente e diventa piú luminoso con l'età. La forza centrifuga di rotazione piú che la pressione sostiene questo ipotetico oggetto supermassiccio e ne impedisce il collasso. Quando viene estratta l'energia gravitazionale, lo spinar si contrae; ne risulta una rotazione piú veloce. Quando una frazione notevole della massa è stata irraggiata come energia lo spinar si contrarrebbe fino a quando la velocità di fuga alla superficie è prossima alla velocità della luce. Collasserebbe poi in un buco nero.

Philip Morrison dell'Istituto di Tecnologia del Massachusetts fu il primo a suggerire che la variazione di luce quasi regolare del quasar 3C 345 può stabilire un'analogia diretta tra quasar e pulsar, con la differenza che nei quasar la scala dei tempi delle variazioni è dell'ordine di mesi invece che di secondi. Se un oggetto supermassiccio avesse un campo magnetico superficiale di 10^5 o 10^6 gauss, le tensioni magnetiche trasporterebbero via l'energia rotazionale a un tasso sufficiente per rappresentare la sorgente di energia di un quasar. Inoltre i campi elettromagnetici variabili indotti dalla rotazione accelererebbero le particelle a velocità prossime a quella della luce. Queste particelle sarebbero responsabili dell'abbondante emissione alle varie lunghezze d'onda: un oggetto supermassiccio rotante di questo tipo si potrebbe chiamare « spinar ». Ci sarebbe una differenza sostanziale tra un pulsar e uno spinar. In un pulsar la rotazione non contribuisce in modo apprezzabile all'equilibrio dell'oggetto, mentre in uno spinar supermassiccio le forze centrifughe forniscono il principale contributo che si oppone al collasso dell'oggetto. Ne risulta un incremento nella rotazione dell'oggetto e un aumento della sua energia rotazionale. Perciò, mentre un pulsar si indebolisce invecchiando, uno spinar diventa piú luminoso. Una complicazione sorge dalla possibilità che una stella supermassiccia, a cui manca la rigidità di una stella di neutroni, non ruoti come un corpo rigido. Se parti diverse dell'oggetto ruotano con velocità diverse, le linee di forza del campo magnetico associato saranno attorcigliate e amplificate. Molti studiosi hanno proposto che tale campo possa esplodere verso l'esterno come una molla e questo processo potrebbe costituire la spiegazione delle radiosorgenti doppie (*si veda la figura a pagina 103*).

Un oggetto in rapida rotazione sarà inevitabilmente molto appiattito per forza centrifuga. Per questa ragione l'evoluzione e la stabilità di dischi rotanti hanno recentemente attirato la attenzione dei teorici. I dischi sono soggetti a vari tipi di instabilità; un disco supermassiccio può frantumarsi in uno sciame di dischi piú piccoli o oggetti simili a pulsar legati tra loro per forza di gravità, ma tutti radianti indipendentemente. Le stelle supermassiccie, i dischi e i pulsar sono in ogni caso piuttosto simili nel comportamento elettromagnetico. In questo senso i due modelli proposti per i nuclei delle galassie – un singolo oggetto supermassiccio o molti pulsar – possono rappresentare stadi diversi dello stesso tipo di sistema.

Come abbiamo visto, una stella supermassiccia in rotazione diventa progressivamente piú luminosa mano a mano che si contrae: è chiaro che questo processo non può continuare all'infinito. Alla fine una frazione notevole della massa dello spinar sarà irradiata come energia. Le dimensioni dell'oggetto saranno tanto ridotte da avvicinarsi al cosiddetto raggio di Schwarzschild, il raggio per cui l'attrazione gravitazionale dell'oggetto è cosí forte che la velocità di fuga delle particelle è equivalente alla velocità della luce. A questo punto diventa decisivo l'effetto della teoria della relatività generale. La pressione interna di un oggetto la cui massa supera notevolmente la massa solare non può opporsi alla gravità dopo l'esaurimento delle sorgenti di calore disponibili. Si ha cosí l'impressione che un oggetto supermassiccio che abbia perduto la maggior parte del momento della quantità di moto collassi e formi quello che si chiama un buco nero. Si può anche pensare che uno sciame di pulsar possa poi agglomerarsi e subire la stessa fine.

Queste considerazioni suggeriscono la esistenza di buchi neri assai massicci nel centro di molte galassie. In realtà, poiché si pensa che la durata dell'attività violenta vari tra un centesimo e un decimillesimo dell'età di una galassia, è possibile che tutti i nuclei galattici abbiano attraversato in qualche momento uno stadio di attività violenta e che vi possa essere un buco nero massiccio persino nel centro di una galassia in quiete come la nostra.

Questi buchi neri aggregano materia dal gas circostante. Se il gas ha un certo momento della quantità di moto, l'aumento non può avvenire lungo linee rette verso il centro del buco nero. La materia tende a orbitare intorno al disco centrale e a formare un disco sottile. Le parti piú esterne del disco ruoteranno piú lentamente di quelle piú interne. La frizione viscosa tende perciò a rompere le parti interne e a trasportare all'esterno il momento della quantità di moto. Via via che il gas della parte interna perde momento della quantità di moto, gira verso l'interno avvicinandosi al centro. In questo modo la materia si muove a spirale verso l'interno con una velocità controllata dall'intensità delle forze viscose fino a essere inghiottita dal buco nero. La viscosità dissiperebbe energia in calore, campi magnetici e particelle veloci: il disco cioè irraggerebbe. L'energia che proviene dall'energia gravitazionale liberata quando la materia cade verso l'interno può ammontare fino al 40 per cento dell'energia totale della massa a riposo della materia accumulata. Si è tentato di interpretare la radiazione non stellare di alcuni nuclei galattici in base a questo tipo di disco. L'aumento di materia da parte dei buchi neri costituisce il meccanismo piú efficiente finora proposto per convertire massa a riposo in radiazione, a parte certi processi strani come l'annichilazione totale della materia quando viene in contatto con l'antimateria.

Tutti i meccanismi descritti implicano strettamente la rotazione. In tutti i casi sembra che sia la rotazione a determinare con quale velocità un sistema può utilizzare la sua energia gravitazionale e l'influenza della rotazione tende a amplificarsi quando un sistema si contrae. Tenendo conto della ovvia importanza della rotazione nella determinazione delle caratteristiche e dell'evoluzione degli spettacolari sistemi su vasta scala quali le galassie a spirale, ci sarebbe da stupirsi se la rotazione non continuasse a essere preponderante anche nei sistemi compatti. Anche nell'ipotesi che alcuni processi particolari su cui ora si indaga non siano pertinenti, sembra però che le idee generali offrano la migliore speranza di comprendere i fenomeni dell'astrofisica delle alte energie, pur mantenendosi entro gli schemi della fisica convenzionale.

GIOCHI MATEMATICI

di Martin Gardner

I « grafi graziosi » di Solomon Golomb, ovvero: come numerare un grafo con parsimonia

Uno dei campi meno esplorati della matematica moderna è costituito da un gruppo di problemi in cui si combinano assieme teoria dei grafi e aritmetica. In questo numero prendiamo in considerazione una categoria di problemi, basati sulla numerazione dei grafi, che è stata recentemente definita e sviluppata da S. W. Golomb, professore di matematica all'Università della California meridionale.

Golomb ha coniato il termine di « grafo grazioso » per definire un grafo che possa essere « graziosamente numerato ». Per spiegare l'espressione egli ricorre a un semplice esempio: la « graziosa »

pare nella figura in basso a sinistra. La figura viene definita « grafo completo per quattro punti », dato che tutti e quattro i nodi di cui è composta sono collegati a coppie da linee chiamate lati. Un grafo del genere è topologicamente equivalente allo scheletro di un tetraedro; esso è inoltre planare, perché può essere tracciato su una superficie piana senza che i lati si intersechino. Come vedremo, un grafo non deve necessariamente essere planare per poter essere numerato graziosamente: unica condizione è che non presenti occhielli (cioè linee che congiungono un nodo a se stesso) o lati multipli (coppie di nodi collegate da più di un lato).

Ogni nodo viene individuato mediante un numero intero positivo. Per convenzione, il numero intero più basso è 0, e non possono esserci due numeri interi uguali. Una volta effettuata la numerazione dei nodi, ogni lato viene individuato mediante la differenza tra i numeri corrispondenti ai due nodi fra cui è compreso. Esattamente come i numeri che definiscono i nodi, anche quelli dei lati non possono essere uguali. L'obiettivo è quello di compiere tutte queste operazioni in modo che il numero più alto fra quelli assegnati ai nodi sia il minimo possibile. Ovviamente, tale numero non può essere inferiore al numero di lati che il grafo presenta. Se il massimo numero di nodo è uguale al numero dei lati *e*, i numeri assegnati ai lati saranno gli interi consecutivi da 0 a *e*, e avremo così ottenuto una « graziosa » numerazione. Il numero *e* rappresenterà allo stesso tempo tre valori: il numero totale dei lati, il massimo numero di nodo e il massimo numero di lato. Qualunque grafo che possa essere graziosamente numerato è un grafo grazioso. Alcuni grafi graziosi presentano un'unica numerazione fondamentale possibile, altri ne hanno più di una (le banali variazioni ottenute mediante operazioni simmetriche quali rotazioni e riflessioni, oppure sostituendo ogni numero *n* di nodo con *e* – *n*, non vengono considerate come diverse). Un grafo che non possa essere numerato graziosamente si definisce grafo sgraziato.

Come osserva Golomb, ogni grafo completo può essere tracciato disponendo tutti i nodi su un'unica retta e aggiungendo i lati rimanenti sotto forma di linee curve (*si veda la parte sinistra della figura in basso a destra*). Andando ancora più in là, immaginiamo che la linea retta sia il margine di un regolo di lunghezza uguale al massimo numero di lati di un grafo numerato. I nodi del grafo sono rappresentati dalle tacche del regolo nei punti corrispondenti ai relativi numeri, in modo che ogni numero indichi la distanza della tacca dall'estremità zero del regolo. Come dice Golomb, un regolo di questo genere è il « modello euclideo » di un grafo completo numerato. Il problema di numerare graziosamente un grafo completo di *n* nodi equivale al problema di tracciare *n* tacche su un regolo (comprendendo come tacche anche le due estremità) in modo tale che ogni distanza fra coppie di tacche sia espressa da un singolo numero intero positivo. Nel nostro esempio il regolo è segnato nei punti 0, 1, 4 e 6, che rappresentano i numeri di nodo di un grafo completo per quattro punti dopo che esso è stato graziosamente numerato. È evidente che un simile regolo può misurare lunghezze di numerazione del grafo che ap- una, due, tre, quattro, cinque e sei unità. Nella figura, alla destra del regolo, è illustrato un altro modo di disegnare il grafo completo per quattro punti: un poligono quadrilatero con tutte le diagonali. Naturalmente l'intersezione delle diagonali non è un nodo. Si osservi che le distanze tra le tacche adiacenti del regolo, insieme alla lunghezza del regolo stesso, corrispondono ai numeri perimetrali del grafo quadrato graziosamente contrassegnato.

In fatto di regoli, un problema del tutto analogo ma meno ristretto viene preso in considerazione nel sesto capitolo del mio libro *La numerologia del dottor Matrix* (Simon e Schuster, 1967). I regoli del dottor Matrix misurano tutte le distanze intere comprese fra zero e la lunghezza del regolo, ma i numeri che contraddistingono i « lati » (cioè le distanze fra ciascuna coppia di tacche) possono anche non essere diversi. Aggiungendo la condizione che siano diversi, il problema del regolo del dottor Matrix diventa identico a quello che stiamo esaminando, e cioè quello di trovare un regolo le cui tacche corrispondano alla numerazione graziosa di un grafo completo con *n* nodi. Golomb dimostra nel suo articolo che il problema può essere risolto solo per *n* = 1, 2, 3 e 4; in altri termini, nessun grafo per *n* punti può essere grazioso se *n* è maggiore di 4.

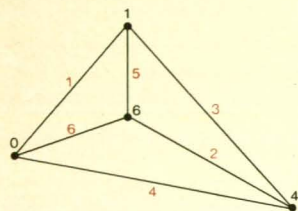
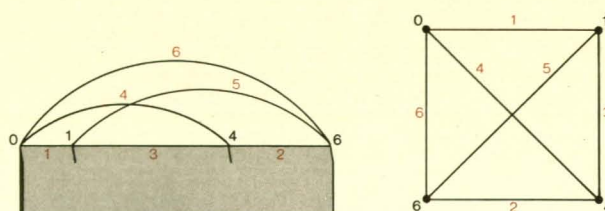
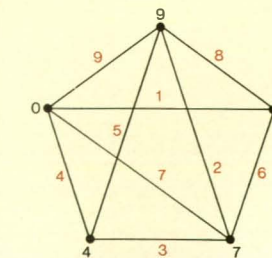
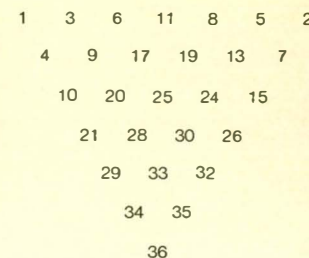
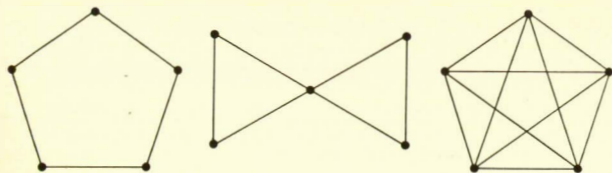
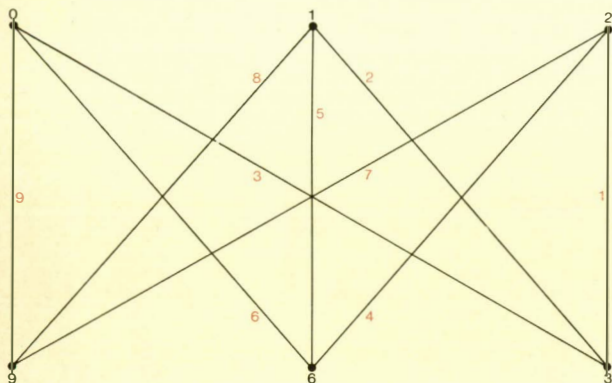
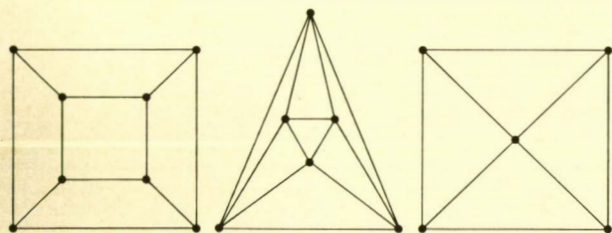
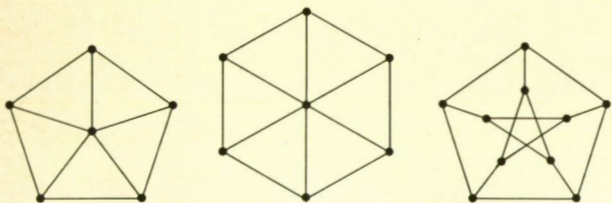
Mantenendo la condizione che tutte le distanze fra coppie di tacche siano diverse, ma consentendo che esse non siano espresse dai numeri consecutivi compresi fra zero e la lunghezza totale del regolo, è ancora possibile cercare di stabilire quale sia il regolo di *n* tacche più corto possibile (sempre comprendendo le estremità come tacche) sul quale tutte le distanze fra coppie di tacche (corrispondenti ai numeri di lato del grafo completo per *n* punti) siano diverse. Nella tabella dei regoli minimi per *n* compreso fra 2 e 11 (*si veda la figura in alto*) soltanto le prime tre righe costituiscono una soluzione al problema del dottor Matrix; esse corrispondono alla numerazione graziosa dei grafi completi per due, tre e quattro punti. Negli altri regoli, le distanze non sono espresse dai numeri interi consecutivi compresi fra zero e la lunghezza del regolo stesso; essi rappresentano quella che Golomb chiama la « migliore » numerazione dei grafi completi per più di quattro punti. I numeri di ogni riga esprimono le distanze fra le tacche adiacenti sui regoli con due, tre, quattro, ..., undici tacche. La tabella, che prosegue all'infinito, si chiama triangolo di Golomb.

NODI	LATI	DISTANZA TRA TACCHE ADIACENTI										LUNGH.
2	1	1										1
3	3	1	2									3
4	6	1	3	2								6
5	10	1	3	5	2							11
6	15	1	3	6	5	2						17
7	21	1	3	6	8	5	2					25
8	28	1	3	6	11	8	5	2				36
9	36	1	3	12	10	8	6	5	2			47
10	45	1	3	6	12	16	11	8	5	2		64
11	55	1	8	10	5	7	21	4	2	11	3	72

TRIANGOLO DI GOLOMB

I minimi regoli di Golomb con un numero di tacche fra 2 e 11.

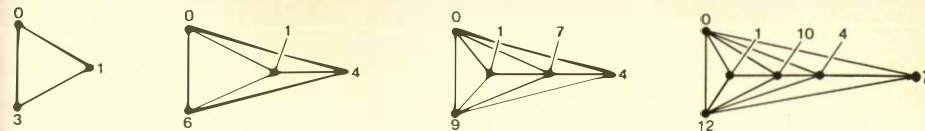
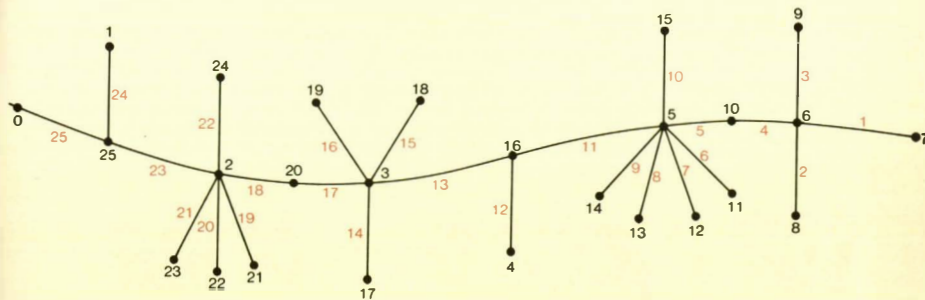
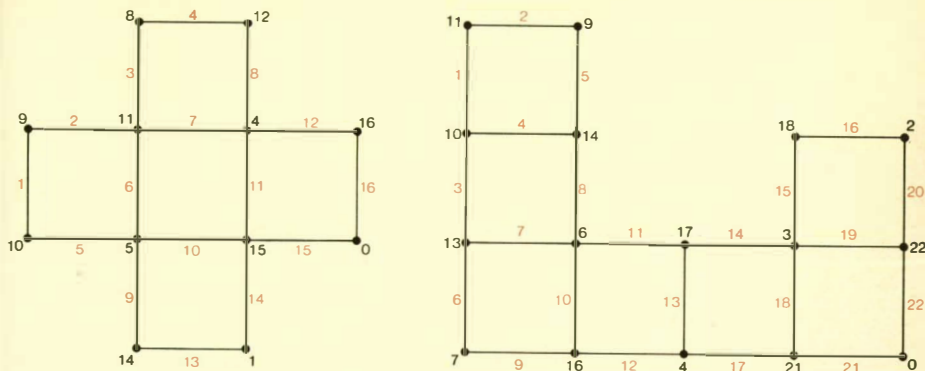
La differenza tra i regoli del dottor Matrix e i regoli di Golomb può essere definita come segue. I regoli del dottor Matrix minimizzano il numero di tacche per ogni regolo di lunghezza *k* che misuri tutte le distanze intere da 1 a *k*; con i regoli di Golomb, invece, per un determinato regolo con un certo numero di tacche, la lunghezza del regolo stesso viene minimizzata, e tutte le distanze intere che esso misura sono espresse da numeri diversi; non è detto, però, che il regolo di Golomb comprenda tutte le distanze intere consecutive da 1 a *k*. Se tracciamo un grafo corrispondente a uno dei regoli del dottor Matrix possiamo anche trovare due lati contrassegnati dallo stesso numero; trascurando tutti i lati con numerazione doppia si ottiene un grafo grazioso che Golomb chiama una « graziosa approssimazione » del grafo completo corrispondente. Per esempio, eliminando un lato (la linea fra i punti 1 e 4) da un grafo completo per 5 punti (*si veda la figura in basso a sinistra*), è possibile numerare graziosamente il grafo stesso. Il risultato è in tal caso equivalente al regolo del dott. Matrix contrassegnato da tacche nei punti 0, 1, 4, 7 e 9.

Prendendo come esempio un regolo di trentasei unità, ecco come si può rapidamente dimostrare che tutte le distanze misurate dal regolo di Golomb sono diverse. Il campione da noi scelto è il più corto regolo noto da otto tacche; la prima riga della figura in basso a destra, ricavata dal triangolo di Golomb, dà le distanze fra le tacche adiacenti del regolo. Questi sette numeri, insieme alla lunghezza totale del regolo, corrispondono ai numeri degli otto lati di un poligono ottagonale che sia stato trasformato in grafo completo tracciandone tutte le diagonali, e che sia stato quindi numerato quanto più graziosamente possibile. La seconda riga è ottenuta sommando le coppie consecutive di numeri della riga precedente; la terza riga si ottiene sommando i numeri della prima a gruppi successivi di tre, la quarta sommandoli a gruppi successivi di quattro, e così via. Il numero in basso rappresenta la lunghezza del regolo, ed è ovviamente la somma di tutti i numeri della prima riga. I 28 numeri che costituiscono il

Un grafo grazioso.

La versione sul regolo (a sinistra) di un grafo completo (a destra).

Grafo grazioso corrispondente a un regolo del dott. Matrix.

La prova del regolo di Golomb a otto tacche.

Gli unici grafi sgraziati con meno di sei nodi.

Numerazione graziosa del grafo di Thomsen.

Tre grafi graziosi di Schlegel: il cubo (a sinistra), l'ottaedro (al centro) e la Grande Piramide (a destra).

Tre grafi graziosi di Golomb con sei, sette e dieci nodi.

triangolo corrispondono ai numeri dei 28 lati del grafo completo per otto punti cui sia stata data la migliore numerazione (sgraziata) possibile. Se i numeri sono tutti diversi, nel grafo completo non esisteranno coppie di numeri uguali, né coppie uguali di distanze fra tacche nel regolo corrispondente.

Golomb ammette che per tutti i regoli di lunghezza superiore a sei unità i risultati sono stati in parte ottenuti (da lui stesso e da altri) mediante procedimento per tentativi; non è stato ancora dimostrato, comunque, che tali regoli siano proprio quelli di lunghezza minima. Il regolo di lunghezza 47, con nove tacche, è stato scoperto per la prima volta nel 1965 da Matthew J. C. Hodgart di Brighton, in Inghilterra; quello di lunghezza 72, con 11 tacche, da Robert Reid di Miraflores, in Argentina, sempre nel '65. Forse qualche lettore sarà in grado di migliorare ulteriormente tali risultati, o di estendere il triangolo verso il basso.

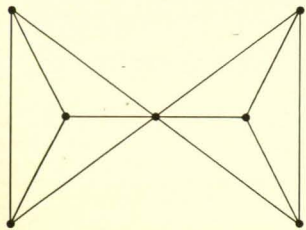
Una delle molte insolite prorietà dei grafi graziosi scoperte da Golomb è quella per cui i loro nodi possono sempre essere divisi in due gruppi – quelli con numero pari e quelli con numero dispari – con il risultato che il numero dei lati che collegano i due gruppi è $[(e + 1)/2]$, in cui e rappresenta il numero totale dei lati del grafo. Le parentesi quadre indicano che l'espressione deve essere arrotondata al numero intero immediatamente inferiore. Questa particolare operazione è definita da Golomb « numerazione binaria ». Per esempio, il gruppo pari dei nodi del grafo a sinistra a pagina 106 è numerato con 0, 4 e 6, mentre il gruppo dispari ha solo il numero 1. Esaminando il grafo, si constata che i due gruppi sono effettivamente collegati da $[(6 + 1)/2] = 3$ lati.

Inoltre, come Golomb dimostra, se tutti i nodi di un grafo sono di ordine pari (con numero pari di lati), il grafo in questione è grazioso solo se $[(e + 1)/2]$ è pari. Quando tale valore è dispari, la numerazione binaria è impossibile e di conseguenza il grafo non può essere contrassegnato graziosamente. Dei 29 grafi topologicamente distinti con quattro o meno nodi, soltanto tre sono sgraziati. Tutti e tre hanno cinque nodi, e tutti i loro nodi sono di ordine pari. Essi violano la legge di Golomb per cui $[(e + 1)/2]$ deve essere pari (*si veda la figura in alto a sinistra*). Si noti che i primi due grafi sono planari, mentre il terzo, cioè il grafo completo per cinque punti, non lo è. Questo dimostra che non tutti i grafi planari e non tutti quelli non planari sono graziosi. È possibile, comunque, che un grafo non planare sia grazioso?

Famiglia infinita di grafi graziosi.

Grafi graziosi a forma di bruco.

I « poliomini » graziosi.

Sì, come dimostra la numerazione del grafo di Thomsen (*si veda la seconda figura dall'alto a fronte*). Il grafo di Thomsen rappresenta schematicamente il famoso (e insolubile) quesito in cui tre case devono essere collegate ciascuna a tre pozzi senza che i sentieri di collegamento si incrocino. Tale grafo fa parte dell'infinita famiglia dei cosiddetti « grafi bipartiti completi », in cui ciascun nodo di un gruppo di a nodi è unito a ciascun nodo di un altro gruppo di b nodi senza però che i nodi del medesimo gruppo siano uniti fra loro. Golomb ha stabilito che tutti i grafi bipartiti completi sono graziosi.

Gli scheletri dei poliedri possono essere rappresentati come grafi planari noti sotto il nome di diagrammi di Schlegel. Dei cinque solidi di Platone, solo il dodecaedro e l'icosaedro corrispondono a grafi per i quali non è stata dimostrata la proprietà di essere graziosi. Abbiamo già visto come si può numerare graziosamente il tetraedro; il lettore è ora invitato a cercare di numerare graziosamente i diagrammi di Schlegel corrispondenti al cubo e all'ottaedro (*si veda la terza figura dall'alto a fronte*), prima di leggere nel prossimo numero la numerazione che a essi assegna Golomb. Il lettore è inoltre invitato a fare altrettanto per il diagramma dello scheletro della Grande Piramide di Cheope, e a scoprire una possibile numerazione graziosa del dodecaedro e dell'icosaedro.

Altri tre grafi graziosi di Golomb hanno sei, sette e dieci nodi (*si veda la figura in basso a fronte*).

Oltre ai grafi bipartiti completi, vi sono infinite altre famiglie di grafi gra-

Un curioso grafo sgraziato.

ziosi. Una di esse, scoperta da Golomb, è quella della prima figura a pagina 109. Sorge una domanda: quando il numero dei nodi tende all'infinito, il rapporto fra i grafi graziosi e tutti i grafi di *n* nodi tende a un limite determinato? E, in caso affermativo, qual è questo limite? Per molti anni non si era potuto escludere alcun valore frazionario tra 0 e 1, ma recentemente Paul Erdös è riuscito a dimostrare che il limite è 0. La prova da lui trovata è però molto complessa.

Sebbene molti dei problemi insoluti sui grafi graziosi, tra cui alcuni di natura decisamente tecnica, siano stati ora chiariti da Golomb, Erdös e altri, sono ancora numerose le domande senza risposta:

1. Quali sono le condizioni necessarie e sufficienti perché un grafo sia grazioso? Non si sa neppure se tutti i grafi ad albero sono graziosi.

Nel 1963, Gerhard Ringel è stato probabilmente il primo ad avanzare la ipotesi, indipendentemente da Golomb e servendosi di una terminologia diversa, che tutti i grafi ad albero possano essere numerati graziosamente. Su questo argomento hanno scritto molto Alexander Rosa e altri matematici cecoslovacchi. L'ipotesi è stata avanzata solo per certi tipi di grafi alberiformi, cioè quelli « a bruco », costituiti da alberi in cui tutti i nodi si trovano sul tronco centrale o alla distanza massima di un lato dal tronco stesso (*si veda la figura al centro a pagina 109*). In un « bruco » tipico, numerato graziosamente, i numeri dei lati si sviluppano consecutivamente da un'estremità all'altra dell'albero.

Golomb ha scoperto un algoritmo analogo per la numerazione graziosa di una classe infinita di « poliomini » del tipo di quelli illustrati nella figura in basso a pagina 109, che sono formati rispettivamente da cinque (pentomino) e sette elementi (ettomino). Si osservi come i numeri si sviluppano consecutivamente in senso diagonale e verso l'alto, da sinistra a destra. Purtroppo esiste una classe infinita di tali grafi, dotati di un maggior grado di concavità (cosa non facile da definire), per i quali il procedimento non vale anche se i grafi stessi possono essere numerati graziosamente.

Un grafo piuttosto semplice, trovato da Golomb (*si veda la figura in alto a fronte*), è particolarmente sgraziato, non potendo essere sottomesso ad alcun teorema generale conosciuto.

2. Quali sono le regole di formazione del triangolo di Golomb? In altri termini, esiste un algoritmo generale con cui si possano trovare i regoli piú corti corrispondenti alla migliore numerazione sgraziata di un grafo completo per piú di quattro punti?

3. Esiste un grafo che, numerato il piú graziosamente possibile, violi l'ipotesi secondo cui in tutti i grafi di tal genere il numero massimo di nodo e il numero massimo di lato sono uguali? Golomb cerca attualmente un controesempio, cioè un grafo che presenti la miglior numerazione possibile, ma in cui il numero massimo di nodo superi il numero massimo di lato (il contrario non è possibile). « Se ne troverò uno – scrive Golomb in una sua lettera – esso sarà non solo sgraziato ma addirittura *disgraziato* ».

Soluzione al problema della piegatura ad anello.

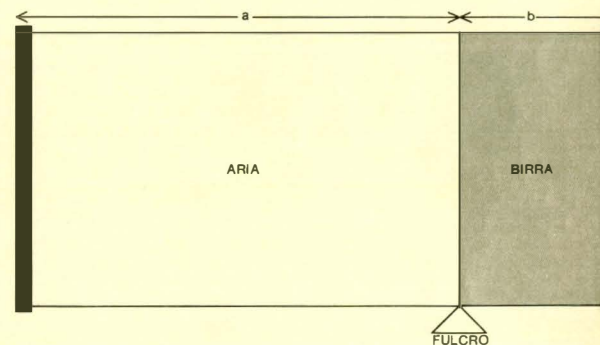
Ecco le risposte ai nove problemi brevi presentati il mese scorso.

1. La figura in basso a fronte mostra come è possibile realizzare un anello con l'estremità *A* attaccata a un tavolo.

2. Le prime due righe della tabella in alto in questa pagina mostrano i risultati delle due prime rotazioni. La prima era stata data il mese scorso; della seconda avevamo detto solo che, dopo di essa, *D* aveva il totale massimo e ciò può appunto accadere solo se i risultati sono quelli della seconda riga. E qui viene il trucco. Sul disco, ogni coppia di cifre opposte ha per somma 5, il che significa che ogni rotazione porta un totale di 5 punti a ogni coppia di giocatori che si fronteggiano, cioè *AD*, *BE* e *CF*. Alla fine del gioco, che ha cinque rotazioni, ciascuna di queste coppie di giocatori avrà un punteggio totale di 25.

Sappiamo che *A* vince il gioco, per cui *D* (che gli sta di fronte) finirà con il punteggio minimo, certamente inferiore a 13. Ma inferiore anche a 12, poiché altrimenti uno dei componenti delle altre due coppie potrebbe avere piú punti di *A*. *D* non può raggiungere quindi piú di 11 e, avendo 9 dopo il secondo turno, farà zero in almeno uno dei turni rimanenti, supponiamo nel terzo (l'ordine delle giocate rimanenti infatti non influenza il risultato); sono cosí determinati i punteggi degli altri due. Nei due turni successivi i punteggi di *D* possono essere 0-0, 0-1 o 0-2, vediamoli. Nei casi 0-0 e 0-2, il punteggio finale di *A* sarebbe ugua-

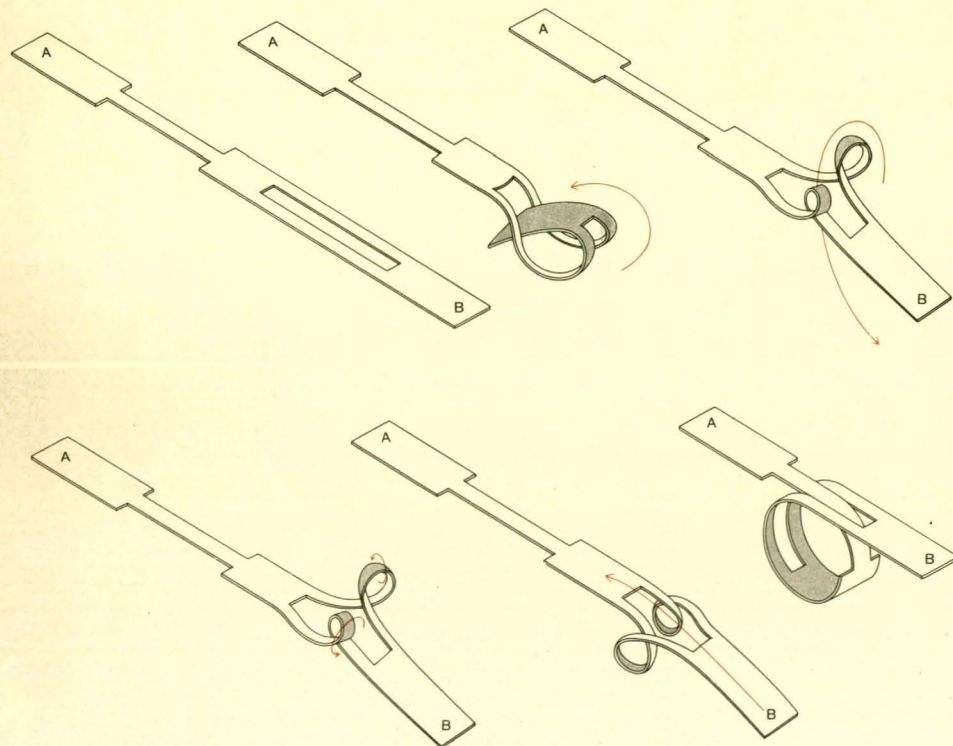
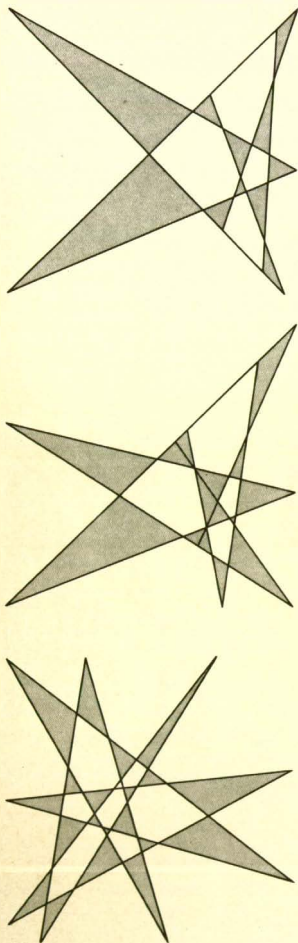
ROTAZIONI	A	B	C	D	E	F
1	1	2	5	4	3	0
2	0	1	2	5	4	3
3	5	4	3	0	1	2
4	5	4	3	0	1	2
5	4	3	0	1	2	5
PUNTEGGIO FINALE	15	14	13	10	11	12

Soluzione al problema del disco rotante.

La lattina di birra in equilibrio.

	75	50	25	0	
VINCE	75	75	75	75	SOMMA = 300
PERDE	−100	−100	−100	0	SOMMA = 300

	125	100	25	0	
VINCE	−50	−50	125	125	SOMMA = 150
PERDE	−50	−50	−50	0	SOMMA = 150

	150	100	50	0	
VINCE	−25	−25	−25	150	SOMMA = 75
PERDE	−25	−25	−25	0	SOMMA = 75

Matrici delle vincite per il giocatore con il dollaro (in alto), con il mezzo dollaro (al centro) e con il quarto di dollaro (in basso).

Soluzione al problema dei triangoli.

le a quello di un altro giocatore. Nel caso 1-1, F farebbe 5-5 e vincerebbe con 15 punti. Resta quindi solo 0-1, il che fa di A il vincitore con 15. Naturalmente non possiamo conoscere l'ordine delle ultime tre giocate.

3. La regola di formazione per il fregio di Coxeter è che i quattro numeri adiacenti

$$\begin{matrix} & b & \\ a & & d \\ & c & \end{matrix}$$

soddisfino l'equazione $ad = bc + 1$.

4. Supponiamo che la birra nella lattina sia congelata per cui la lattina possa essere messa orizzontalmente in equilibrio sulla lama di un coltello con la parte alta verso sinistra. Se si ha equilibrio con il fulcro sotto la parte piena, l'aggiunta di birra la farà pendere a sinistra, la sottrazione di birra a destra; se all'equilibrio il fulcro è sotto la parte vuota accade il contrario. Se però all'equilibrio il fulcro è proprio in corrispondenza alla superficie della birra, *qualsiasi* cambiamento nella quantità di birra la farà pendere a sinistra (*si veda la figura al centro a pagina 111*). Poiché in questo caso il centro di gravità si muove verso l'alto della lattina per qualsiasi variazione della quantità di birra, il centro di gravità sarà nel punto piú basso quando coincide con la superficie della birra.

Con la lattina cosí in equilibrio immaginiamo di eliminare le estremità e di distribuire la loro massa sulla superficie della lattina. Questo non altera l'equilibrio perché non sposta il centro di gravità del sistema, ma ci permette di considerare la lattina come un tubo aperto la cui massa per unità di lunghezza della parte vuota (sinistra) è proporzionale al peso di una lattina vuota e la cui massa della parte piena (destra) è proporzionale al peso di una lattina piena. Il momento della forza a sinistra è quindi proporzionale al peso di una lattina vuota moltiplicato per il quadrato della lunghezza della porzione vuota (sinistra), mentre il momento della forza a destra è analogamente proporzionale al peso di una lattina piena moltiplicato per il quadrato della lunghezza della porzione piena (destra). Poiché siamo all'equilibrio, i due momenti devono essere uguali e quindi, con ovvie notazioni, $a^2V = b^2P$, cioè $a/b = \sqrt{P/V}$. Con i dati forniti il mese scorso si ha quindi $a = 12$ e $b = 4$; il centro di gravità si trova nel suo punto piú basso quando la lattina è piena per un quarto.

5. Qualunque sia la scelta della moneta il gioco è equo. Le matrici delle vincite (*si veda la figura in basso a pagina 111*) dimostrano questa asserzione a prima vista incredibile.

6. Il numero massimo di triangoli non sovrapponentisi che si possono costruire con 7, 8 e 9 segmenti di linea sono rispettivamente 11, 15 e 21 (*si veda la figura in questa pagina*).

7. $532 \times 14 = 98 \times 76 = 7448$.

8. Se n è dispari non vi sono soluzioni. Se n è pari ma non multiplo di 4 all'ultima mossa occorre saltare un numero dispari di pedine, il che necessariamente lascia alla fine una pedina singola nella riga. La tesi che vi è una soluzione quando n non è multiplo di 4 è quindi falsa.

Se vi sono $4n$ pedine il problema può essere risolto ragionando a ritroso in base al seguente algoritmo. Si parta con $n/2$ dame in una riga. Si prenda la pedina superiore da una delle due dame centrali, la si muova verso il gruppo maggiore di dame e la si disponga da sola. Nella successiva mossa all'indietro si prenda la pedina superiore dell'altra dama centrale e la si muova nella stessa direzione fino a superare la pedina sola. Si segua questo procedimento finché non sono eliminate tutte le dame nella direzione del primo salto. Si prenda ora la pedina superiore della dama piú interna e la si muova nella stessa direzione precedente; si continui questo procedimento, sempre nella stessa direzione finché tutte le dame non sono ridotte a pedine. Eseguendo queste mosse in ordine inverso si ottiene una soluzione del problema (ve ne possono essere altre).

9. Ecco sei possibili soluzioni al problema delle traccie degli sci.

1) Lo sciatore si è appoggiato all'albero. Mantenendo uno sci fermo ha ruotato accuratamente l'altro sci attorno all'albero; ritornato nella posizione iniziale, ha sollevato il piede utilizzato, su un solo sci è passato dall'altra parte dell'albero, ha riposato lo sci sulla traccia lasciata in precedenza e ha ripreso la discesa.

2) Lo sciatore si è spiaccicato contro l'albero mentre gli sci hanno proseguito la corsa.

3) Gli sciatori erano due, ciascuno con un solo sci.

4) Un solo sciatore ha percorso la discesa due volte, ogni volta con un solo sci.

5) Uno sciatore ha percorso una discesa senza alberi allargando una sola volta gli sci. In quel punto è stato posto subito dopo un albero senza radici.

6) Si trattava di uno sciatore con gambe abbastanza lunghe e larghe da passare sopra l'albero.

I problemi pubblicati nel numero di gennaio hanno attirato l'attenzione dei lettori. Moltissimi hanno contestato infatti la soluzione del problema 21 come incompleta. Hanno perfettamente ragione. Oltre alla soluzione $n = 0$ il problema ammette infatti la soluzione $n = 1000$. Uno di questi lettori, il signor V. E. Orlando di Roma, ci ha anche segnalato due altre soluzioni al problema 12 sottolineando, giustamente, che forse ce ne sono altre ancora. Eccole:

AL – LI – IN – NE = LIANE
FA – AL – LI – IO = ALFIO.

Ringraziamo tutti quanti ci hanno scritto mandandoci le loro soluzioni.

LIBRI

Che cosa ci direbbe la Beagle *oggi?*

di Liana Bolis

Nel 1835 la *Beagle* raggiunse l'isola di Chatham nelle Galapagos: la piú orientale del gruppo.

Charles Darwin si trovò davanti a uno spettacolo strano, fantastico: sabbie di lava nera, quasi nulla di verde: alberi scheletrici, rocce, tutto sembrava bruciato, mentre serpeggiavano stanche lucertole di una specie ignota tra le orribili rocce.

Non sembrava certo di essere arrivati nella quiete del Pacifico, alle tanto decantate Galapagos, le isole cosí diverse da tutte le altre isole della Terra.

Ma ben altra era la realtà delle Galapagos, e ben altra l'impressione che esse dovevano fare su Darwin.

Nella baia di Santo Stefano la natura offriva una serie di nuovi animali: tartarughe giganti, pesci tropicali, squali e la sabbia era ancora nera e calda. Particolarmente interessante fu proprio l'incontro con le enormi tartarughe: gusci di tartarughe erano disseminati lungo la spiaggia e ciò indicava la grande strage che gli indigeni ne facevano: iguana e lucertole marine godevano della libertà di isole lasciate a sé, di cui ben poco l'uomo si era interessato. Salendo piú in su nell'interno dell'isola, l'aria umida aveva splendidamente favorito la natura e gli alberi erano ricoperti di felci, di orchidee, di licheni. All'isola di James, Darwin trovò ben 26 specie di uccelli, mai visti in altre parti del mondo.

Egli notò che tutte queste isole anche poco distanti tra di loro possedevano una serie di uccelli molto diversi, e tale differenza egli la notò particolarmente nei fringuelli; un piumaggio dal nero lava al verde di varie tonalità a seconda del luogo. Nessun uccello, a eccezione del picchio dal becco scarlatto, aveva i colori vistosi della zona equatoriale. Diversissimi i becchi dei fringuelli: tale differenza fu ben facile vedere in rapporto alle possibilità di trovarsi il cibo. Becchi robusti per spezzare semi e nocciuoli; becchi sottili e snelli per nutrirsi di teneri insetti, fiori e frutti.

Cosí in queste varie isole, tanto diverse per vegetazione e sviluppo di vita, i fringuelli si erano adattati alle diverse condizioni per sopravvivere.

Era dunque possibile che l'ambiente tanto avesse influenzato sulla specie? Era possibile che l'isolamento avesse favorito lo svilupparsi di nuove specie?

Questo era il quesito che si poneva Darwin in quell'anno, di fronte a tanta diversità offertagli dall'osservazione dei fenomeni naturali in un'area cosí piccola.

Charles era partito quasi quattro anni prima come naturalista a bordo della *Beagle*, un brigantino messo a disposizione dall'Ammiragliato inglese per fare dei rilievi geologici e naturalistici in una crociera intorno al mondo al comando del capitano FitzRoy. Apparteneva a una famiglia della buona borghesia di Cambridge e suo padre lo avrebbe voluto vedere laureato in medicina come egli stesso; ma alla fine, viste le reticenze di Charles, lo aveva destinato alla carriera ecclesiastica. Il contrasto, quindi, tra la regola di vita e di pensiero alla quale era stato destinato, con quello che la sua mente stava elaborando era grandissimo: in piú FitzRoy, l'aristocratico comandante, con il quale aveva modestamente discusso i suoi pensieri, era un uomo estremamente legato agli scritti della Bibbia, e certamente era sconvolto dall'idea che lo stesso libro della Genesi non fosse null'altro che una leggenda.

Ma Darwin aveva già notato, osservato e assorbito nel profondo dei suoi pensieri le diversità della natura nel suo viaggio fin lí; di carattere timido, fragile di salute, sofferente del mare, continuava a raccogliere esemplari nuovi che mandava regolarmente in Inghilterra; a ogni tappa faceva lunghi percorsi nel retroterra.

Il suo primo incontro con la foresta tropicale fu molto importante, sia per lo spettacolo offertogli dalla maestà della natura, sia per l'osservazione della crudele legge della foresta tra gli animali. Lo aveva particolarmente colpito la lotta spietata di una vespa con un ragno, la marcia delle formiche soldato, le mimetizzazioni operate da tante specie per sopravvivere. E non meno, lungo il viaggio, lo aveva colpito la cruda realtà della violenza umana, nella guerriglia al confine meridionale dell'Argentina; il terribile terremoto del Cile; l'acqua che dal profondo dell'oceano saliva altissima e trascinava via tutto quanto.

Alla Terra del Fuoco avevano sbarcato tre fuegini, due uomini e una donna che FitzRoy aveva portato in Inghilterra in un precedente viaggio. Essi sembravano perfettamente adattati al comportamento, ai modi e alla nuova lingua inglese; ma quando Darwin li vide dopo poco tempo dallo sbarco nel loro paese, essi sembravano voler dimenticare quel loro scorcio di vita trascorso nella civiltà. Avevano lasciato i vestiti europei, che con tanta ambizione avevano indossati, avevano dimenticato l'inglese, e fieri erano ritornati alla loro vita primitiva che aveva dato loro la vera sensazione dell'uomo legato alla terra, figlio dell'universo: e forse avevano inconsciamente sentito quello che avrebbero potuto perdere. FitzRoy era particolarmente toccato da questo fatto: il suo carattere duro e aristocratico, dell'uomo che osserva la natura senza vivere in essa e con essa, era tanto diverso da quello di Darwin.

L'Australia alle coste perfettamente inglese, linda e curata, era tanto diversa dalla vera Australia dell'interno: gli aborigeni, splendida popolazione legata alla terra e che segue la sua terra vivendovi intimamente nelle lunghe marce per trovare cibo e ansietà di vita, seguendo affascinanti rituali, cominciavano a sentirsi a disagio perché continuamente indotti ad « accettare la civiltà ». Ancora oggi i pochi aborigeni sono aiutati ad accettare i *settlements*, mentre la loro vita naturale è una realtà che non si può cancellare.

Alle isole Cocos, Darwin sente ancora urgente il suo desiderio di pensare alle modificazioni geologiche e di vegetazione in rapporto agli animali marini e terrestri; l'evoluzione di atolli, le barriere coralline e le scogliere davano la reale impressione della instabilità della Terra.

Darwin trovò l'isola Mauritius un vero paradiso; e ancor oggi lo è: una natura calda e offerente con acque limpide e con una fauna piú che affascinante.

DARWIN E LA BEAGLE di Alan Moorehead, Rizzoli, Milano, 1972 (L. 8000).